



Article

Tackling Verification and Validation Techniques to Evaluate Cyber Situational Awareness Capabilities

Salvador Llopis Sanchez ^{1,*}, David Sandoval Rodriguez-Bermejo ^{2,3,4}, Roumen Daton Medenou ^{2,5}, Ramis Pasqual de Riquelme ², Francesco Torelli ⁶ and Jorge Maestre Vidal ^{2,*}

¹ Communications Department, Universitat Politècnica de Valencia, 46022 Valencia, Spain

² Indra, Digital Labs, 28108 Alcobendas, Spain; david.sandoval@tarlogic.com (D.S.R.-B.); rdaton@indra.es (R.D.M.); gramis@indra.es (R.P.d.R.)

³ Department of Computer Science, University Carlos III de Madrid (UC3M), 28005 Madrid, Spain

⁴ Tarlogic, 28050 Madrid, Spain

⁵ Computer Science Department, University Internacional de La Rioja (UNIR), 26006 Logroño, Spain

⁶ Leonardo-Finmeccanica, 00195 Rome, Italy; francesco.torelli@leonardo.com

* Correspondence: salllosa@masters.upv.es (S.L.S.); jmaestre@indra.es (J.M.V.)

Abstract: Since cyberspace was identified as a domain of operations, defence practitioners started a race with academy, researchers, and industry and military organizations working together towards defining related lines of capability development (e.g., DOTMLPFI) and exploring the needs and opportunities they entail. An essential cornerstone of adapting to the convergence of the cyber domain with conventional theaters of operation is the need for producing tools for easing to acquire cyber situational awareness (CSA), from which human operators shall be able to perceive, reason and project situations and events observed in cyberspace that may vertically/horizontally propagate from technological to tactical, operational and strategic planes. Benefiting from the higher maturity level of civilian capabilities for cybersecurity, the military sector has embraced the challenge of creating related beyond state-of-the-art CSA enablers that comprise the existing technological background while adopting concepts such as operations, missions or courses of action (CoAs), properly aligning them with military doctrine. Beyond ongoing development efforts, there is a wide methodological gap in the lack of suitable CSA verification and validation (V&V) frameworks, which are expected to analyze if related capabilities meet the requirements to operate in the military context; at the same time supporting the thorough development life-cycle of brand new cyber defence technologies. With the motivation of closing the identified gap, this research introduces a novel V&V framework able to guide the evaluation of CSA-related tools, which makes converge purely military aspects with dual-use state-of-the-art V&V approaches. Three core CSA evaluation concepts are discussed in-depth: software, operational and application tests. They range from the daily application of new capabilities to their ability to enable the acquisition of a joint operational picture understandable by human decision makers.

Keywords: cyber defence; cyber situational awareness; decision making; verification and validation

MSC: 28E99; 68M15; 90B25; 91C99



Citation: Llopis Sanchez, S.; Sandoval Rodriguez-Bermejo, D.; Daton Medenou, R.; Pasqual de Riquelme, R.; Torelli, F.; Maestre Vidal, J. Tackling Verification and Validation Techniques to Evaluate Cyber Situational Awareness Capabilities. *Mathematics* **2022**, *10*, 2617. <https://doi.org/10.3390/math10152617>

Academic Editor: Todor Tagarev

Received: 6 May 2022

Accepted: 21 July 2022

Published: 27 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The concept of *situational awareness* (SA) is colloquially known as “knowing what’s going on so you can figure out what to do” [1,2], more recently redefined by M. Endsley as a human state of mind that allows decision makers to “perceive the elements in the environment within a volume of time and space, to comprehend their meaning and project their status in the near future” [3] (see Figure 1). The digitalization of the defence sector and the consolidation of cyberspace as a fifth domain of operations has led to the need for understanding the impact of perceived situations on the cyber defence infrastructure, services

and assets on traditional physical domains, resulting in the evolution of conventional SA to the understanding of the convergence between cyberspace and other dimensions, referred to as cyber situational awareness (CSA) [4]. However, despite its relevance, today, no single solution fulfills all operational requirements needed for a comprehensive solution tailored to the military, highlighting, among other things, the importance of advanced information filtering to adjust SA pictures to match the needs of each particular user (strategic, operational and tactical), reduce information overload and increase operational efficiency. This is due, among other things, to the great disruption of cyber operations at a conventional doctrine level, the difficulty of making converge the mission concept with the technological background inherited from the dual ecosystem, and the information technology dynamism inherent in the rapid evolution of the digital sectors. On the other hand, and despite there being a wide interest of cyber defence practitioners in acquiring CSA-related capabilities, their validation entails a *rara avis* topic in the cyber defence landscape [5], which, contradictorily, implies an outstanding gap in the thorough design and development of CSA-related tools with particular needs for adapting to emerging communication paradigms [6,7].

Motivated by these changes and with the purpose of contributing to digitalizing the defense sector, the research presented in this paper describes a joint effort between military industry, academy, research and technology organisations towards defining a common evaluation methodology by using V&V techniques able to assess the capability of cyber defense tools to guide the human acquisition of CSA, which is expected to assist their development and maintenance cycle, while assessing their effectiveness prior to consideration for deployment on real theaters of operation. The conducted research supports the findings previously presented in [8]. Its main contributions are enumerated as follows:

- In-depth review of the state-of-the-art research in military CSA, as well as the evaluation methodologies for assessing the cyber defence tools for CSA acquisition, highlighting needs, technological gaps, challenges and opportunities.
- Introducing a core set of evaluation concepts and a testing workflow for validating CSA related tools.
- Proposing novel verification and validation guidelines for mission-centric assessing of the vertical/horizontal propagation of cyber threats between cyberspace and the mission plane.
- Describing CSA tool-evaluation guidelines and suggesting a reference questionnaire template.

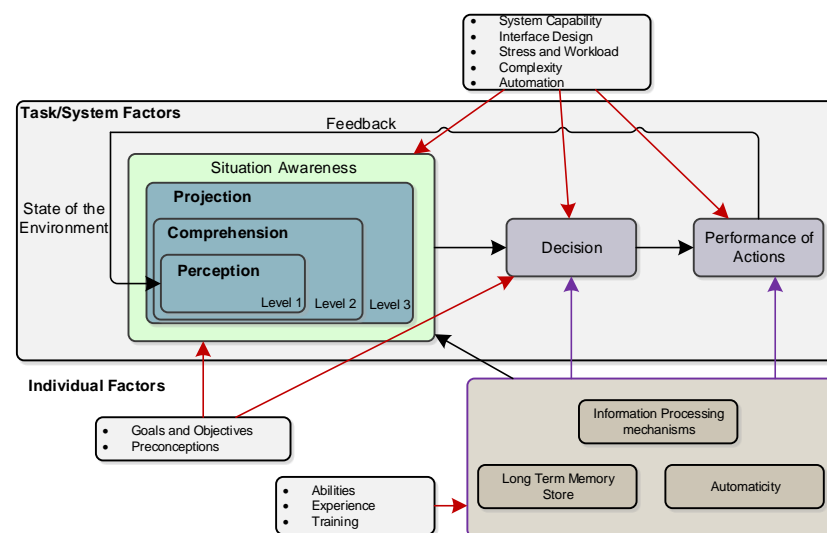


Figure 1. Endsley’s model for situational-awareness acquisition [9].

The paper is organized into eight sections, starting with an introduction. Section 2 reviews the state-of-the-art research in CSA and discusses the main challenges and method-

ological gaps of the related evaluation procedures. Section 3 describes the cornerstones of the introduced framework for evaluating CSA-related tools. Section 4 presents the testing concepts of the framework. Section 5 presents its operational concepts, and Section 6 presents its application concepts and details the proposed evaluation workflow. Finally, Section 7 summarizes the conclusions and potential lines of future research.

2. A Review of The State-of-the-Art Research

The following describes the evolution and emerging challenges in the fields of SA, CSA in military operations and the assessment of CSA-related tools.

2.1. Situational Awareness in Cyber-Physical Systems

Endsley proposed three stages of information processing: perception, comprehension, and projection. The first stage entails tasks such as monitoring and incident identification, the second stage involves their analysis and association, and the third stage forecasts the development of the system's state. These should support security operators' decision making by defining a closed-loop strategy that implements continuous feedback between the inferred operational picture, decision making (course-of-action identification, selection, planification) and their enforcement. The Endsley model's adaptation has shown to be especially successful in complex and dynamic cyber environments [10,11], where the definition of "cyber situational awareness" was observed; the diagnosis is strongly dependent on the context in which occurrences are recorded, and where CSA becomes crucial in the battle against cybercrime. Alternatively to the Endsley's model, the acquisition of SA has been widely studied and adapted during the last decades. For example, a generalized variation has emerged as an outstanding solution for enhancing the effectiveness of first-response actuation (Figure 2): the observe–orient–decide–act (OODA) loop. Originally conceived for supporting decision making in uncertain and chaotic environments, the OODA loop adapts the scientific method for solving countermeasure identification, selection planning and enforcement challenges; where observe resembles the acquisition of preliminarily factual knowledge, orient–decide proposes the best hypothetical approach to a cyber-physical incident, and act coincides with testing and contrasting the assumed hypothesis. If, based on the next observations, the planned response seems improvable, the decision loops are executed again looking for a better solution. This procedure allows responding to complex situations in near real time, and increasing the effectiveness of the first-response actuations during their course. However, implementing this loop in heterogeneous technological ecosystems, as is the case of real operations, requires the development of an accurate operational picture, which shall be provided by combined cyber-physical awareness acquisition [12]. For example:

- **Observe.** Perception of the operational environment through the aggregation and fusing of data from diverse data sources, directly collected by sensors watching over cyberspace or physical features, or gathered from social-cognitive sources (media, etc.) and cyber-physical intelligence repositories or entities.
- **Orient.** The deduction of additional knowledge useful for the in-depth understanding of the operational, where diagnosis, prediction, simulation and related adaptive machine-learning capabilities shall allow identifying cyber, physical and combined threats; as well as anticipating their evolution at different time horizons.
- **Decide.** Supports decision-making and planning/conducting first-response actuations on the identified threats by taking advantage of the analytical capabilities previously adopted at the Orient step of the OODA loop. The suggested decisions will provide predictions and simulations able to define anticipatory actuations and preliminarily assess the impact of the courses of action to be conducted.
- **Act.** Reactive/proactive responses to be conducted will take advantage of both technological enablers that take part of the protected tunnel infrastructure and those carried by first responders. Consequently, conventional actuators such as fire extinguishers, ventilation systems, signalling devices, etc., shall be combined with

next-generation capabilities, e.g., unmanned vehicles, wearable devices or virtual network functions (VNF).

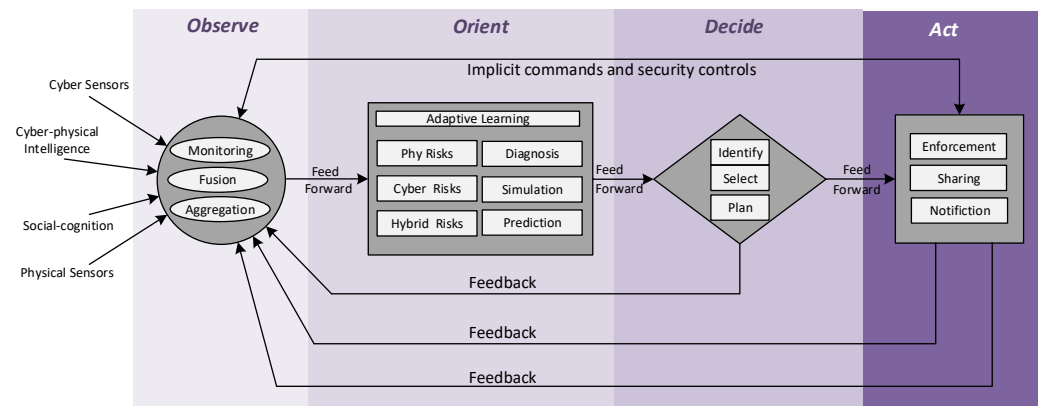


Figure 2. Example of OODA loop for assisting cyber-physical incidents.

This procedure allows a response to complex situations in near real-time and improve the effectiveness of the first-response actuations during their course [12]. Many of the related preliminary contributions were collected in [13], where an in-depth revision of the CSA landscape evidenced the predominance of approaches towards facilitating emergency risk management, industrial systems, and communication networks [14]. As discussed in [10], they enhance the information security risk management (ISRM) process's three most prevalent flaws: (1) information-security risk identification is frequently performed in a superficial manner; (2) information security risks are frequently estimated without much consideration of the current situation; and (3) information-security risk assessment is frequently performed on an ad-hoc, non-historical basis (a conventional security risk assessment scheme can only provide a "snapshot" of cyber risks at a given time).

Numerous publications attempt to combine both paradigms in order to address these issues while maintaining the ISRM foundation. This is the case of [15–17], where the SA is acquired, taking into account the definition of risks, assets, and their influence on preventing, identifying, and mitigating cyber threats presented by various standards and platforms. In [11,18,19], this is extended to emerging communication environments, where to share the acquired operational pictures constitutes a key element for inferring cyber threat intelligence (CTI) [20] and cyber-attack attribution. In [21,22], the problem of sharing acquired operational pictures was discussed, where a method for defining critical information and the relevant information quality elements that are required to build shared SA were discussed. On these grounds, the OODA loop became a key enabler for near real-time acquisition of SA in critical scenarios [12], where the responses teams must adapt their modus operandi as their mission progresses.

In the meantime, security practitioners have witnessed an exponential growth in the development and deployment of various types of cyber-physical systems (CPS) [23] able to cover insider/outsider sourced situations [24], which can be found in a plethora of areas, such as aerospace, automotive, chemical processes or transportation, most of them being tagged as critical Infrastructure. However, as highlighted in this study, one fundamental issue in CPS security is the heterogeneity of the building blocks—which implement hardware components such as sensors, actuators, and embedded systems—and in the grounds of different collections of software products, proprietary and commercial, for control, monitoring and any kind of tasks. As a result, every component, as well as their integration, can be a contributing factor to a CPS risk, in this way presenting a wide variety of risk surfaces [25]. The complexity of CPS and the heterogeneity of its components have encouraged the proposal of different approaches towards addressing their safety [26], and its combination with security and privacy protection [12], which have a close relationship with CSA military needs.

As the complexity of physical interactions increases, social factors became fundamental for the proper management and understanding of cyber-physical situational awareness enablers [22] and prevent safety and security situations [23]. Consequently, in [25], the need for a paradigm shift from classical CPS to cyber-physical social systems (CPSS) was anticipated, where it was notorious for analysing the impact of CPS on humans and vice versa, hybridisation, i.e., machines and human users covering parts of the system function in deep interaction, is emerging as a novel core concept. Therefore, CPSS can be thought of as a new step in the development of ubiquitous computing, which typically entails three stages: CPS, cyber-social systems (CSS), and CPSS [26]. These two systems, in particular, together with CPS and CSS, are improving and becoming universally interconnected. In order to facilitate intelligent interaction between the cyber, physical, and social realms, they have finally been integrated into CPSS, in this way presenting a perfect baseline for developing capabilities for managing hybrid threats from operations in the *grey zone* and/or hybrid warfare theaters of operation.

2.2. Managing Cyber Incidences on Military Operations

The CSA on military operations extends the scope of the conventional CPS security capabilities towards covering the understanding of the dependences between the cyberspace and the operational context, which includes the rest of the battle domains (air, land, space, sea) and the ongoing/planned missions [27]. The concept measure of effectiveness (MOE), applied to mission impact assessment and hostile adversarial actions in cyberspace, was introduced by Musman [28], and refers to the measures that characterize the operational effectiveness of a cyber defence unit or force in achieving its objectives during a joint mission [29]. The authors explicitly stated the need to determine the mission impact of a cyberattack based on the timing and duration of the attack. This was first noted by Fortson in his thesis in 2007 [30]. According to [28], the timing/duration of the attack needs to be correlated with timing and workflow information from the mission in order to precisely infer consequences on MOE. For that purpose, mission models should capture such time-dependent information of a mission, including the relative importance of its various tasks along the time. Some of the state-of-the-art mission-mapping technologies that include dependencies with cyberspace were analysed in [31], being classified as process-driven analysis techniques or artifact-driven analysis techniques. Overall, the study concluded that the process-driven acquisition of CSA [32] was typically developed manually by subject-matter experts, which evidenced drawbacks that are expected to be partially addressed by the inclusion of automatism: (1) they are highly subjective; (2) they are highly time-consuming; and (3) mission mapping cannot be updated in a timely manner. On the other hand, artifact-driven analysis techniques [33] are close to the data fusion [34] paradigm, where logs and other data collected by sensors shall be able to guide the discovery and/or inference of cyber assets, their dependence regarding planned/ongoing missions and potentially vertical threat propagations. Compared to process-driven approaches, these techniques are much faster, support real-time asset discovery and can identify non-trivial dependencies. However, they heavily rely on capabilities for monitoring, discovering and analyzing cyberspace, as well as their capability of making the information that they provide understandable by operators and decision-makers.

On the other hand, when acquiring CSA, it is critical to understand the key assets that contribute the most to supporting mission processes and services [35]. These assets are commonly referred to as cyber key terrains (CKT). In a recent work, Price et al. explore the use of multi-attribute decision-making (MADM) for identifying them [30]. An interesting observation given by Price et al. is that the most important factor in determining the CKT is the relative importance of each subsystem for a particular mission. This is valuable, since the number of subsystems can be manageable for a process-driven approach but the number of assets (where an asset is part of one or more subsystems) is certainly not. In [35], the authors argue that a complete understanding of the current cyber operational picture is not necessary to effectively manage the impact of cyber threats. As an alternative

to holistic solutions, they propose Cyber-ARGUS, which leverages the contributions by Damico et al. [36] for the construction of an ontology system to simplify mission impact-assessment tasks.

Based on the analysis of existing technologies and publications and the feedback received by stakeholders regarding previous publications, the authors of this paper can conclude that none of the related works provides a complete set of required functionalities at an adequate maturity level for supporting military operations, where the following knowledge areas require further research: active data gathering to construct more reliable mission mappings, identification of mission dependencies in non-recurring missions, timely updating of mission mapping as mission needs evolve, and security implications for the mission if an attacker compromised the mission mapping. As a result, the research community should at least prioritize the following developments: (1) asset perturbation to probe and measure mission sensitivity to every network asset; (2) mission-aware architectures, which heavily rely on software-defined networking and virtualisation techniques to deploy system architectures that already reflect the mission dependencies; (3) YMAL (you may also like) approaches based on the machine learning and big data of past missions to help the analyst during the mapping of capabilities to the current mission needs; (4) and a role-based mission-behaviour baseline that correlates the role of the user executing the tasks in the mission planning system and the assets supporting their functions.

2.3. Situational Awareness Assessment

To truly understand “what is going on” in cyber defense, it is rarely simple to obtain a sense of the threat landscape as a whole [9]. Endsley’s definition of situational awareness defined progressive levels of awareness, ranging from (1) simple perception of fundamental data, (2) interpretation and combination of data into new knowledge, and (3) the capability of forecasting and anticipating situations and courses of action (CoA). This shows that situational awareness can be attained gradually, rather than in a linear fashion, with greater degrees of understanding dependent to some extent on lower levels of awareness [37]. This preliminarily conception settled the ground for developing a plethora of more or less elaborated and validated methods for measuring situational awareness on physical operational environments [2], particularly focused on the air battle domain. However, as pointed out by Brynielsson [38], to determine whether it is possible to evaluate/test a cyber solution in terms of achieved cyber situational awareness heavily relies on the capability of CSA measurement based on identifying suitable SA activities transposable to the cyber domain, and those inherent in the cyber domain that are not significant in the rest of the battlegrounds. Depending on the necessity, this activity might involve a small-scale validation test or a full-scale CDX utilizing a design that allows for relevant training while also assessing the degree to which the cyber solution has contributed to individual comprehension of the broader cyber situation.

Bearing this in mind, the objective of the measurement shall be to be able to compare an object or event with another. With this purpose, Stevens stated that, for measurement, it is essential that “numbers are assigned to aspects of objects or events according to one or another rule or convention”, thus allowing to quantitatively conduct comparisons to, ideally, an object of truth in order to be able to rate the level of SA [39], with this being a well-known statement heavily supported by the research community. It is worth highlighting that the concept of truth or “ground truth” in the SA context has been an object of controversy for decades, where authors such as Parasuraman argue in favor of its existence [40], and others such as Dekker criticizing the feasibility of acquiring it; the latter was justified by the need for an objectivity [41]. To deal with this problem, the research community has developed and validated different techniques that often describe models particularized to their application domain. A good example of this is illustrated in [42], where a large inventory of situational awareness methodologies for C4I (command, control, communications, computers and intelligence) environments applied to domains such as military aviation, air traffic control, energy plants, etc., were presented. It divided the

situational awareness methods into seven great categories: (1) SA requirement analysis, (2) freeze probe analysis, (3) real-time probe techniques, (4) self-rating techniques, (5) observer rating techniques, (6) performance measures, and (7) process indices. They are briefly described, as follows:

- **SA requirement analysis.** Requirement analysis will identify the components that commanders acquire in situation awareness in specific operational contexts and domains. This analysis is thought of as the preliminary action for all SA assessment actions. It may also involve questionnaires, goal-directed task analysis, and interviews with subject-matter experts [43].
- **Freeze probe analysis.** Freeze probe analysis is based on freezing or temporally interrupting the testing activities for conducting direct queries to participants regarding their understanding of the operational domain. The responses are contrasted with a sequential ground truth, which allows quantitative/qualitative measurement of the acquired operational picture. The main benefit of this procedure is its direct nature, thus allowing gathering fresh feedback from the operational domain. However, the freezes may disrupt the conducted tasks, which may alter the forthcoming results. The situation awareness global assessment technique (SAGAT) [44] has been consolidated as one of the most effective freeze probe analysis techniques, and is used in many different contexts, such as military aviation, air traffic control, military operations, driving, industrial processes, etc. [45].
- **Real-time probe techniques.** In contrast with the freeze probing analysis methods, real-time probe techniques query users without interrupting their tasks. For this purpose, non-intrusive questionnaires must be preliminarily developed by subject-matter experts, which indicate the topics to be considered, and the time in which questions should be conducted. There is controversy of the “non-intrusively” property of this approach, since most queries may direct attract the participants’ attention and potentially bias the data. The the situation present assessment method (SPAM) is the most widely applied and adapted real-time probe technique [46], which was preliminarily conceived for assessing the situational awareness acquired by air traffic controllers.
- **Self-rating techniques.** As a variation of the aforementioned methods, self-rating techniques assess the commander understanding of the operational picture through a rating scale. These questionnaires are typically conducted post-tests, so they are not intrusive regarding the tasks to be conducted. As well as not disturbing the participants during the test execution, they are easy to administer. However, they are criticized from two perspectives: first, since the questionnaires are not conducted online, they may entail poor recall and bypass sensitivity/stress variables. On the other hand, the participants may misunderstand the situation or tamper with the self-rating scores. The situation awareness rating technique (SART) is the best self-rating method in the state bibliography [3], which is administered post-trial. SART has 10 dimensions, each of which is rated on a seven-point scale.
- **Observer rating techniques.** The most popular field rating methods use observers who are typically subject-matter specialists. The fact that these methods are non-intrusive and adaptable to real-world settings is their fundamental benefit. It is debatable if the subject-matter expert can grade the learned situation awareness appropriately. For example, within this category, one extensively used observer evaluation method for infantry men is the situation awareness behavioral rating scale (SABARS) [47]. This method consists of 28 observed behaviors that are scored on a five-point scale.
- **Performance measures.** Situation awareness can be measured indirectly using performance indicators. These metrics can range from hits and misses to false alarms, accurate rejection, and response time, depending on the task and circumstance. The relationship between performance and situation awareness is not always evident, despite

the fact that performance metrics are frequently simple-to-use and non-intrusive [5]. This is due to the fact that these indicators may be directly provided by automatism.

- **Process indices.** Process indices are based on assessing behavioral and/or biometrical features that may reveal the participant condition when decision making (e.g., stress, confidence, etc.). For example, this is addressed in [48] by eye-tracking on the SA operator's gaze. The primary criticisms of this strategy are supported by the fact that the recorded data only document what participants observed, not what they perceived [49].

In [50], according to Endsley, (good) SA can be viewed as a factor that raises the likelihood of good performance but does not ensure it. Good system design decisions can be made by measuring situation awareness, which eventually increases the likelihood that the operator will make wise judgments and avoid poor ones. She worked to make sure a technique was valid and reliable in order to create effective measurement procedures:

1. Establishing metrics that only measure the object the procedure is meant to assess.
2. Employing sensitivity and diagnostic techniques to provide the necessary understanding.
3. Using a probing technique that is balanced for each specific purpose.
4. During the procedure, the construct should not be altered significantly.

These considerations became particularly relevant when assessing cyber situational awareness, where the acquired picture presents a hybrid nature between human factors and supportive analytical processes conducted by automatisms.

2.4. Verification and Validation of the Acquired Cyber Situational Awareness

One of the pioneering publications for evaluating CSA tools was introduced by Brynielsson et al. [38], and aimed to transpose some of the general-purpose SA assessment methodologies to the cyber domain, which was driven by the execution of cyber defense exercises (CDX) and adapting the questionnaires and criteria previously considered in kinetic contexts. In [5], it is remarked that most of the existing related evaluation efforts were designed to be technology-oriented, thus focusing on assessing the performance and quality of the acceptance of the supportive technological enablers, rather than how the presented operational picture impacts on the participant understanding of the operational environment. Criticized by the research community, this simplified the scope of the CSA solutions into data fusion and artificial knowledge inference tools [51]. In response, Gutzwiller et al. [52] coined the term cyber-cognitive situational awareness (CCSA), which attempted to back the CSA point of view into the human activities conducted in cyberspace. It is important to keep in mind that cognitive factors were originally presented in the classic SA models, so, as highlighted by M. Endsley in [37], the common understanding of cyber situational awareness as a mere result of log analysis or data-driven actions entails a fallacy that may lead to assuming inaccurate CSA design/validation decisions. Although the number of publications regarding the role of the human factor in CSA-related topics have increased ([53–56]), few have presented research on which actual CSA measurement technique has been used [57] and evaluated in realistic scenarios.

Among this research, it is worth highlighting the contributions of Giacobe et al., where a cyber SAGAT questionnaire was conceived for evaluating the participant's acquired CSA [58]. Analogously, in [59], SART and SAGAT were adapted to CDX, where it was concluded that Giacobe's measures were not feasible mainly due to their implementation difficulty. In order to reduce the intrusiveness of the questionnaires into the tasks conducted by the participants, in [58], SPAM was implemented for assessing CSA in the context of collaborative cyber security, which reveals a weak correlation between the acquired CSA and the observed performance. However, some studies, such as [5], oppositely concluded that freeze probing measurement tactics are feasible in particular scenarios, for example, large CDX. In [60], SAGAT was successfully applied in the context of the Pacific Rim Collegiate Cyber Defence Competition (PRCCDC), a CDX held by the U.S. Service Academy, where problems derived from acquiring quality data and modelling each participant status (anxiety, enthusiasm, etc.) thorough the exercises, were considered. The research is about

what should be visualized in log analysts [61], and previous task analyses in the field [62] give some insight into what information is important to be processed for facilitating cyber situational awareness acquisition. To establish operationally pertinent situation awareness, a CSA system must make use of six classes of information (threat environment, anomalous activity, vulnerabilities, key terrain, operational readiness, and ongoing operations). These constitute a strong understanding of CSA as a whole, but they still need to be further defined and examined. Shiravi et al. [61] particularized these outcomes by conducting a review of network security visualizations, identifying the six primary data sources that may be used: network traces, network events, application logs, security events, as well as contexts pertaining to network activities, users, and assets. However, because they only consider anomalous actions and not the other five criteria of Dressler's classification, these data sources only cover a small part of CSA. In addition, these approaches do not consider the vertical propagation of risks between the communications and information systems (CIS) layer and the mission layer, which entails a significant handicap when solely applied in supporting cyber defence operations.

Since the evaluation methodology proposed in this paper attempts to combine both holistic cyber situational awareness assessment and log/event-driven measurements, the following sections consider more deeply the most outstanding evaluation statements considered in the fields of: cyber risk detection and inference, cyber and mission risk assessment and capabilities for supporting military decision-making capabilities.

3. Verification and Validation Frameworks

The following presents the design assumptions and challenges, and the evaluation concepts that integrate the proposed framework.

3.1. Assumptions and Problem Statement

Software (and system) requirement quality control, which includes validation and verification, is frequently described in the literature as a heterogeneous process utilizing a wide range of largely independent methodologies [63]. Since no standard for software requirements validation has been specified for CSA tool development and evaluation, the state of the art has been widely reviewed with the observation of a large variety of heterogeneous classical and novel methods for completeness, consistency, validity, realism, ambiguity, verifiability, etc. and their checking procedures (test case generation, prototyping, requirement review, automated consistency analysis, walk-through, etc.). Inspired by Zimek et al. [64] on meeting the problem of multiple truths in data from clustering and pattern mining perspectives, the Indian parable of the blind men and the elephant perfectly describes the need for an ensemble perspective towards producing a holistic analysis of a CSA-related tool. The parable describes how different "experts", in this case blind men, reach different conclusions about the nature of the animal based only on the information offered by touching a different part of its body. By pooling their knowledge, experts were able to find the correct classification of the creature. In the context of CSA verification and validation, each of the reviewed proposals provided a partial and interesting view of the characteristics of the analysed tool, but it is assumed that only their merging will allow answering questions such as those addressed by Jackson's model [65] for proper assessment of the fulfillment of software requirements and specifications, among them:

- *What is the effect on the environment that is desired?*
- *What would the behaviour of the machine on the boundary (interface) be that would achieve the desired effect on the environment?*
- *What would the internal behaviour and structure of the machine be that would lead to the needed behaviour on the boundary?*

The answer to question (1) led to the definition of the outer requirements of the system. On the other hand, by answering question (2), the inner requirements of the system (which are usually meant when one speaks of "requirements") are defined. Finally, question (3) refers to the design principles of the solution to be developed for satisfying the outer and

inner requirements (see Figure 3); all of them need to be addressed by the evaluation framework to be proposed.

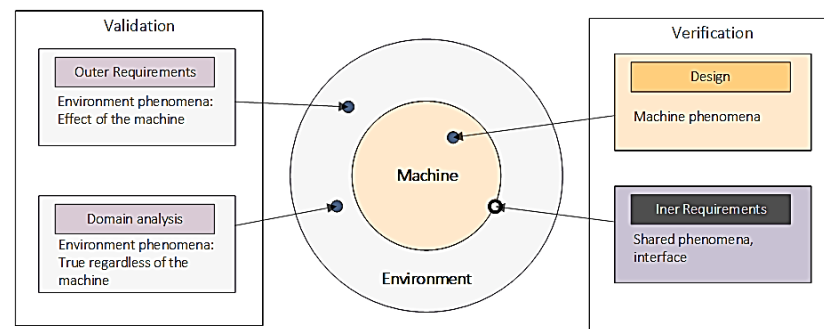


Figure 3. Jackson's model: environment, machine and boundary [65].

Due to this, software end users and customers tend to focus on external needs, whereas designers and developers pay close attention to internal requirements. A more recent classification associates inner requirements to functional capabilities and outer requirements to non-functional features, which shall be separately defined at the design stage of CSA tools. Therefore:

- **Functional requirements**, related to the capability or functions required to directly support users in completing their goals and missions.
- **Non-functional requirements** emerge as system requirements to suit the users' functional needs, and are often implicit and technical in nature. Examples include service quality, availability, timeliness, and accuracy.

Each operational necessity and authoritative source, such as a document or a person, must be attributable to a functional or non-functional requirement. Once defined, it is given a special identification number that makes it possible to precisely connect the need to the software design, coding, and testing processes. The task related with validating and verifying requirements shall be measurable, either quantitatively or qualitatively. On the other hand, quantitative measurements typically refer to numbers. A quantitative approach can be used to count events, metrics or numerical KPIs directly gathered from tests, demonstrations or the operational effectiveness of the solution. The definitions of the requirements must be quantifiable or qualitatively measurable. Typical risk management categories include:

- **Measures of effectiveness (MOEs)**, i.e., metrics of mission success from the perspective of the stakeholders.
- **Measures of performance (MOPs)** used to assess whether the system satisfies the performance standards required to fulfill the MOE.
- **Key performance parameters (KPPs)** or indicators (KPIs) defined by stakeholders as measures of a minimal and critical system or project performance and level of acceptance.

In the context of the previous contents, a clear separation is observable between the concepts validation and verification, which shall answer the following questions:

- **Validation:** *am I building the right product (according to end-user expectations)?*
- **Verification:** *am I building the product correctly?*

System validation activities are typically driven by demonstrating the fulfilment of outer requirements, while system verification activities typically assess inner requirements. Consequently, the term validation can be defined as "the assurance that a product, service, or system meets the needs of the customer and other identified stakeholders"; while verification can be defined as "the evaluation of whether or not a product, service, or system complies with a regulation, requirement, specification, or imposed condition". A holistic evaluation methodology for verifying and validating CSA tools shall strongly analyse the

relationship between both techniques, since, as defined by Endsley, situational awareness depicts a human state of mind that facilitates means and ways of the evaluation of CSA tools. Based on this, the proposed methodology addresses the challenge of introducing novel evaluation concepts and models that, in the grounds of ergonomics, shall guide the definition and execution of verification tests at all levels (technical, operation, application, etc.), perform the validation of non-functional requirements assuming a dual-use end-user perspective, and establish a logical linkage between them so that, at the end, it shall be possible to state if a CSA-related tool is already able to facilitate the user acquisition of cyber situational awareness.

3.2. Evaluation Model and Core Concepts

The evaluation of CSA capabilities, both as subsystems or integrated platforms as a whole, implies a multidisciplinary action that shall be suitable for covering any phase through the life cycle, from early design and implementation actions to their maintenance and update/upgrade once delivered and integrated into end-user facilities. Bearing this in mind, the introduced evaluation methodology distinguishes three main evaluation concepts, which shall be studied in depth by following a bottom-up direction: testing, operation and application; each of them targeting specific evaluation criteria (see Figure 4). As a first step towards evaluating CSA-related systems (from down to top in Figure 4), testing processes will be conducted for validating the developments functionality with an intent to find whether the developed software met the specified requirements or not, and to identify the flaws to ensure that the product is glitch-free in order to produce a quality product. This concept embraces the development of software-unit, integration, security and reliability tests prior to evaluate the targeted capabilities on uses cases, as well as their application in operational environments.

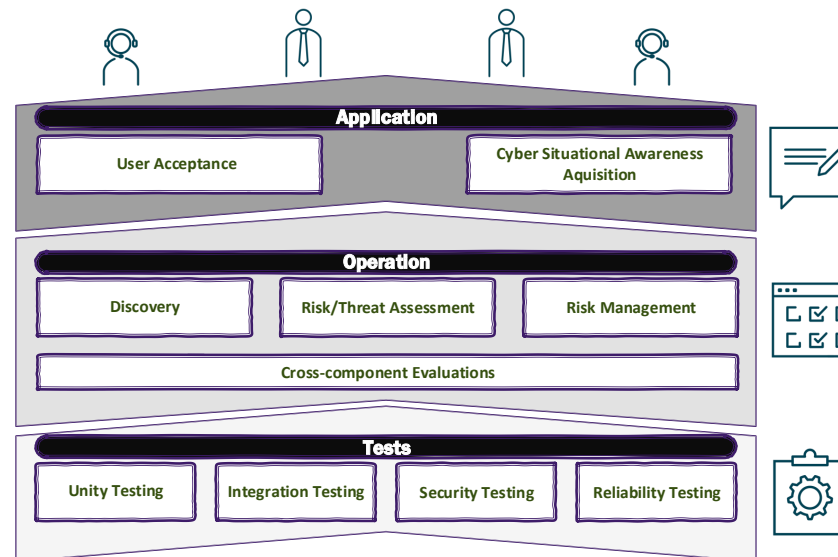


Figure 4. CSA verification and validation framework.

The second step is the evaluation of the targeted capabilities when operating. These actions cover a preliminary cross-component evaluation that includes accuracy, performance, response time, updatability, upgradability, scalability, and strengthening against adversarial tactics. They can be assessed separately (i.e., component-to-component, function-to-function, etc.) or assuming an integrated CSA solution as a whole. Then, a particular evaluation will be conducted, which is focused on the capabilities for mapping the cyberspace, the capabilities for assessing the vertical propagation of threats from the cyberspace to the mission plane, and the capabilities for supporting their management.

Finally, the third step refers to the evaluation of the studied solutions for properly facilitating the acquisition of CSA, which is assumed as their major purpose. This includes evaluation of the end-user response by tracking both biometrical and cognitive features. In addition, the user acceptance will be validated by directly asking end users about their level of satisfaction, as well as their opinion concerning the usability, ease of use, functionality and suitability of the information presentation.

3.3. Limitations of the Proposal and Conducted Research

The early identified and assumed limitations in the scope of the conducted research are:

- The issues of responsible and safe data management are not specifically addressed by the concepts proposed. Therefore, prior to adoption, methods to enforce privacy and data protection shall be studied independently.
- The suggested methodology provides a broad overview but will need to be customized for the unique characteristics of each application scenario. The information presented above should help analysts evaluate and adjust cyber defense capabilities and the operational context in which they will be used.
- Time and resource restrictions have not been taken into account. Therefore, before adoption in situations with limited resources, a suitable frugal adaptation may be required.
- Prior research on the subject of cyber situational awareness from a mission-centric approach is seriously lacking for varied and strong contributions. This indicates that the proposal's theoretical underpinnings are unstable, and, as a result, the proposal may need to be modified when the scientific community in this area of study makes further progress.
- The conducted research skips over potential adversarial tactics that could sabotage the proposed evaluation activities. These may range from direct assaults on sensors that monitor performance indicators to poisoning threats against human-computer interfaces.

3.4. Early Identified Operational Risks

The following anticipated circumstances may make the proposal less likely to be adopted.

- The sensors used, as well as how they communicate internally, may have an impact on the conducted measurements.
- It is possible that users lack the education and/or training necessary to use the cyber defense solutions that are being tested.
- The data gathered during the operation of the cyber defense technologies may report insufficient observations, which may result in misleading statistical conclusions.
- Models built from training data may be used by sensors, data fusion techniques, etc. The research community is aware of the problem of the lack of appropriate datasets in some study domains, which also arises from operational situations such as those to be evaluated by the methods developed.
- The proposed surveys and validations might not take into account the variations in cognition and skill of different users. Therefore, the evaluators' capacity to handle this issue will ultimately determine how effectively the methodology is implemented.
- The user operation may be disrupted by probes, potential intermediary surveys, etc., generating a visibly apparent fake operating context. This may impact the evaluation results.
- The users may behave differently if they are conscious of being monitored.

4. Cyber Situational Awareness Testing Concepts

This methodology describes four testing tasks concerning the proposal testing evaluation concept: unity testing, integration testing, reliability testing and security testing, which may be customizable, updatable or ignored as the capability development plan requires.

4.1. Unit Tests

Unit tests are typically automated tests written and run by software developers to ensure that a section of an application (known as the “unit”) meets its design and behaves as intended [66]. This methodology assumes that units are the smallest testable part of a CSA-related solution, since they enable discovering early problems through the development lifecycle. Moreover, unit testing allows developers and experts of all technical partners and stakeholders to amend, refactor or, if code needs to be fixed later, make sure the module still functions properly. Test cases should be tested independently, taking advantage of methods such as:

- **Method stubs:** they may act as a temporary stand-in for still-to-be-written code, or they may emulate the behavior of existing code.
- **Mock objects:** simulated items that accurately mirror the behavior of real objects. In order to test the behavior of another object, a programmer often builds a mock object.
- **Fakes:** a fake is closer to a real-world implementation than a stub. They can be useful for system tests as well as for unit testing purposes, but they are not intended for production use because of some limitation or quality requirement.
- **Test harnesses:** collection of software and test data configured to test a program unit by running it under varying conditions and monitoring its behaviour and outputs. They are able to: (1) call functions with given parameters, output the results, and (2) compare the obtained values. The test harness serves as a hook to the created code so that it may be tested automatically.

This methodology assumes that units are the smallest testable part of a cyber-situational-awareness-related solution, since they enable discovering early problems thorough the development life-cycle. Moreover, unit testing allows developers and experts of all technical partners and stakeholders to amend, refactor or fix code at a later date, and make sure each module—as a single entity within a modular system configuration—still works correctly. As a code base became larger, an existing set of unit tests simplified integration of new code modules. Integration testing became more simpler by first testing the individual components of a software before testing the programme as a whole.

The definition of the most suitable test will be designed, as illustrated in Figure 5, based on customer requirements or their transpositions, among them those concerning the presentation layer, system workflows, processes, functions, storage, etc. The implementation of the capabilities that allow their fulfilment will be identified at all levels, from framework objects to schema, structs, etc. The meta-design considers the entire application from a holistic point of view and is concerned with such issues as: object decoupling through the use of design patterns, an application-wide framework, componentization of different functional blocks, and instrumentation.

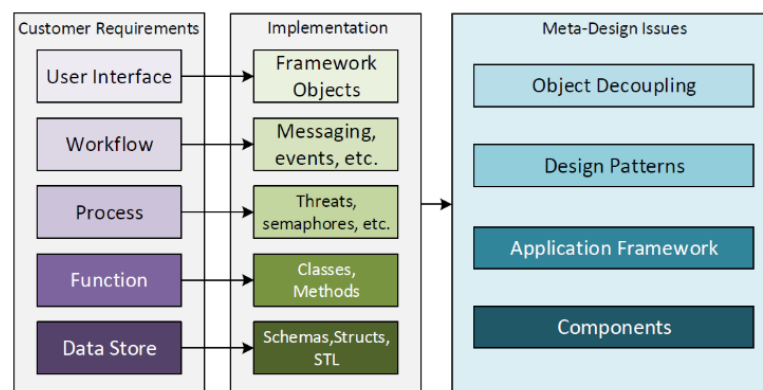


Figure 5. Global unit test requirements.

4.2. Integration Tests

Individual software modules are merged and tested as a unit during integration testing. Unit-tested CSA components are used as input for integration testing, which collects them into bigger aggregates and applies tests specified in an integration test plan to those aggregates to produce an integrated system suitable for final testing [49]. Running user-like workloads in integrated user-like environments serves as the foundation for this kind of integration testing. This method of testing directly proves the environment while indirectly verifying the individual components through use. This methodology heavily relies on the component developers to perform the isolated unit testing for their input on the CSA systems, as planned. The strategy's objective is to avoid repeating the developer testing and instead identify and address issues brought on by the interactions of the components in operational environments and virtual labs.

Integration tests typically may be conducted driven by three approaches: big-bang, bottom-up and top-down, and/or their combination. This methodology assumed that the most suitable approach shall be decided in the scope of the testing needs, considering, among other things, the granularity of the tests, development stage or the nature of the use cases to be defined. The following briefly describes each of them:

- **Big-bang testing:** In order to create a complete software system or a significant portion of the system, the majority of the generated modules are connected together and used for integration testing. The integration-testing procedure can be sped up significantly with the help of this technique. However, the integration process will be made more difficult and the testing team may not be able to accomplish the purpose of integration testing if the test cases and their results are not properly documented.
- **Bottom-up testing:** Prior to using them to facilitate the testing of higher level components, the lowest level components are tested. Up till the component at the top of the hierarchy is tested, the process is repeated, including the integration and testing of all the bottom-level or low-level modules, processes, or functions. The next level of integrated modules will be generated and can be used for integration testing when the lower level integrated modules have undergone integration testing. This strategy only works when all or the majority of the modules at the same stage of development are completed. This approach also makes it simpler to provide testing progress as a percentage and aids in assessing the technology readiness levels of generated applications.
- **Top-down testing:** the top integrated modules are tested and the branch of the module is tested step by step until the end of the related module. This is the opposite of bottom-up testing.
- **Hybrid/sandwich testing:** a mix of top-down and bottom-up strategies. Here, lower modules are integrated with top modules and tested at the same time that top modules are tested with lower modules. Stubs and drivers are both used in this method. The degree of hybridization used and the module scope for each technique determine the pros and cons.

4.3. Security Tests

In order to determine whether a software application is secure or not and for determining whether it is vulnerable to cyberattacks, security testing is a useful method of software testing [67]. Specifically, (1) to determine if an information system protects data while maintaining functionality; (2) to check how the information flows in a secure environment; (3) to assess how the CSA capabilities react when being disrupted; and (4) to identify software weaknesses in terms of confidentiality, integrity, authentication, availability, authorization and non-repudiation. Targeting this direction, this methodology suggests taking advantage of commercial off-the-shelf (COTS) solutions, as is the case of the OWASP (Open Web Application Security Project) Testing Guide [68] as the basis for evaluating the CSA-related developments and the supported services discovering potential flaws, improper configurations, or risky end-user behaviors. Other related COTS solutions are

Nmap/Zmap, Generic online network tools (whois/domaintools, etc.), Burp Pro, Charles and ZAP proxies + Burp extensions, Maltego, IntelliJ, etc. The key steps of the adopted security testing approach are the following:

1. **Discovery.** To better understand the application's scope and functionality, the technologies and design concepts being used, and any attack vectors, the application should be manually walked through.
2. **Configuration management.** Analyse the servers that support the application. This includes any web servers, application servers, database servers, or load balancers that are accessible from the target system. These systems are analysed for missing patches, up-to-date software, and security-related configuration settings.
3. **Authentication.** Ensure that the application properly verifies the user's identity before granting access to restricted functionality and data within the application. Authentication testing also seeks to determine if the authentication process was coded and configured according to recommended best practices.
4. **Authorization.** Ensure the authorization controls by manipulating cookies, hidden parameters, and other identifiers and to attempt to access resources and functionality without having an active authenticated session.
5. **Session management.** Check for issues that may allow a user's session to be hijacked or otherwise compromised to permit an attacker to impersonate the victim within the application.
6. **Data validation.** Ensure the application handles user input and output securely to prevent misinterpreting user input strings as executable commands or database queries. This also applies to the potential impacts of data forgeries and data-driven adversarial machine-learning tactics (e.g., mimicry). In the case of data feeds from perceptions of the operational environment, validations may include procedures to avoid cognitive bias or intentional actions.
7. **Error handling.** All error messages will be checked that are returned for any sensitive or useful information.
8. **Data protection.** Assessment of the effectiveness of sensitive data protection in storage or transit due to a lack of encryption; improper use of production data in a test environment; or displaying sensitive information to an unauthorized user.
9. **Reporting.** Proper documentation of the security test results, ranging from human-understandable texts up to binary, logs, etc. and any other evidence generated.

4.4. Reliability Tests

Reliability in statistics and psychometrics is the overall consistency of a measure, which refers to producing similar results under consistent conditions. In the context of CIS (as is the case of CSA enablers), and as indicated by the American National Standards Institute (ANSI), The chance of fault-free software operation for a certain amount of time in a given environment is what is known as software reliability [69], which became essential when deployed in military operations on the cyberspace. This methodology relies on the following reliability testing procedures, which may be extended as each capability requires: *feature testing*, *regression testing*, and *load testing*. The first covers three actions: (1) each function in the software should be executed at least once; (2) interaction between two or more functions should be reduced; and (3) each function should be properly executed. On the other hand, regression testing is conducted when new functionalities are added or obsolete ones are removed, in order to ensure that consequent issues/bugs were not introduced. Finally, load testing attempts to stress the CSA systems under the maximum workload expected for identifying its limitations.

As frequent in the state of the art, the reliability will be measured according to the mean time between failures (MTBF), which is the anticipated amount of time between inherent failures when the system is operating normally. Accordingly, for each observation, the "down time" is the instantaneous time it went down, which is after (i.e., greater than) the moment it went up, the "up time". The difference ("down time" minus "up time")

is the amount of time it was operating between these two events. With reference to the illustration above, the MTBF of a component is calculated by dividing the total operational time by the total number of observed failures. Accordingly:

$$MTBF = \frac{\sum(\text{startof downtime} - \text{startofuptime})}{\text{TotalNumberofFailures}} \quad (1)$$

5. Cyber Situational Awareness Operational Concepts

The CSA operational concepts include the V&V of capabilities for dynamically assessing risks/threats on the cyberspace, inferring their potential operation and impact on the mission plane, and suggesting the most suitable CoAs. These procedures are described below.

5.1. Assessment of Cyber Threats

In the last decades, different methodologies for evaluating risk/threat recognition systems have been published. They monitor additional factors that aid in determining which detection procedures are most appropriate for each use case in addition to their capacity to detect mismatches. Keep in mind that, as shown in [70], in the same way that risk/threat discovery is affected, the evaluation of an intrusion detection system that only takes accuracy into account gives a snapshot of its effectiveness in a specific moment in time. Therefore, this comparison with future proposals tends to become less relevant when modifications are made to the monitoring scenario. As a result, the research community has also been given additional guidelines for determining their effects on a monitoring environment [71]. Accuracy, effectiveness, responsiveness, ease of updating, scalability, and resilience to evasion attacks are the most widely used in the bibliography. Refs. [72,73] are also adoptable as criteria for the evaluation of the risk/threat identification effectiveness on CSA solutions.

5.1.1. Accuracy

The accuracy of the procedures for identifying risks and threats describes their capabilities for distinguishing neutral, adversarial and friendly situations on the cyber domain, as well as the deduction of mission-level situations from them. These calculations classify timely monitorizations and events observed in the operational environment, labeling them accordingly; typically, as positive (i.e., the situations belong to a particular class) or negative (i.e., the situations do not belong to a particular class). Four subcategories are typically used to evaluate these labels: true positives, true negatives, false positives, and false negatives [74].

Based on the indicators, it is easy to state that the capability objectives for inferring risks and threats from the operational environment is to achieve the highest true positive and negative rates, in the same time that the false positive and negative rates are minimized. On these bases, the considerations in Table 1 are suggested to assess the class-based accuracy of the discovery capabilities in cyber situational awareness:

Despite its simplicity, this binary classification typically fails to evaluate responses in terms of the accuracy of a system regarding variations in its adjustment parameters, where factors such as the tolerance to calibration errors, or the support to solving trade-offs between accuracy and other features (e.g., resource consumption, data protection and privacy, etc.) demand additional measurements. On the other hand, the evaluation of non-binary classification samples also requires the introduction of additional considerations. The first of them is typically addressed by taking advantage of receiver operating characteristic (ROC) analysis, while the second leads to calculations of a confusion matrix.

Table 1. Accuracy: illustrative deconstruction of observation in classes.

Feature	Description	Classes
Analysis for CTI for update /feasibility	Accuracy for CTI assessment when updating/upgrading, and deletion of CTI linked to FPs	Binary: an action may be needed/ non-needed Multiclass: kind of action
Threats/Alerts Grouping	Accuracy when grouping threats and/or alerts	Binary: a threat/alert may be grouped/ non-grouped Multiclass: Particular group/cluster on which they are associated
CI threat/risk detection	Accuracy when identifying CI threats and/or risks	Binary: a CI situation may entail a threat/risk, or not Multiclass: a particular group/cluster on which they are associated
Mission threat/risk detection	Accuracy when identifying MI threats and/or risks	Binary: an MI situation may entail a threat/risk, or not Multiclass: a particular group/cluster on which they are associated
Threat/risk Prediction	Accuracy when forecasting threats/risks in the requested time horizons	Binary: a sequence of observations may derive in a threat/risk at time horizon t [75] Multiclass: the kind of threat/risk from which the sequence may be derived.
CI to MI propagation	Accuracy when inferring MI threats/risks from CI threats/risks	Binary: a CI threat/risk may derive an MI threat/risk, or not Multiclass: the kind of threat/risk from which a CI threat/risk may be derived.
Anomaly recognition	Accuracy when discriminating	Binary: an observation is tagged as discordant, or not Multiclass: the kind of anomaly in which an observation may be tagged.
CKT recognition	Accuracy when recognizing cyber key terrains (CKT)	Binary: a cyber asset is tagged as CKT, or not. Multiclass: the kind of CKT in which a cyber asset is tagged.
Multi-step attack	Attacker next step estimation	Binary: a sequential attack may be derived from further intrusion steps, or not Multiclass: the kind of attack steps inferred from the current status of a multi-step attack.
Attack path evaluation	Accuracy when inferring each attack path	Binary: an attack path is inferable from certain attack, or not Multiclass: the kind of attack path inferred from a certain attack.
Self-protection issues	Non AuthZ/AuthN actions, user inactivity, etc.	Binary: an observation is tagged as a potential self-protection issue, or not Multiclass: the kind of self-protection issues from which a certain situation is derived.

5.1.2. Performance and Efficiency

The efficiency of a cyber situational awareness solution determines its ability to process information as a function of time. On the other hand, performance refers to how the information processes manage to reach a certain efficiency. Although the performance of CIS is widely studied in the state of the art, there are a few contributions concerning efficiency engineering from the software architectural point of view. The reason for the lack of literature could be that efficiency engineering is usually seen as a part of performance engineering. As already mentioned, improving efficiency will boost performance in almost all cases. Nevertheless, efficiency engineering is also about optimizing resource allocation. Therefore, in the following part of the methodology, the performance or efficiency optimization will be the same, but what is different is the requirement (time constrain or resource limits). The following measurable factors may be considered in the context of assessing the performance and efficiency of cyber situational awareness solutions:

- **Latency.** Time delay between the cause and the effect of each process in the system. Latency is a result of the limited velocity with which any system interaction can take place (measured delays, packet jitter, etc.).
- **Bandwidth.** Measurement of the bit-rate of available or consumed data communication resources (Mbit/s, Gbit/s, etc.).
- **Throughput.** Rate of production (Packets/s, Transactions/s, Events/s, etc.)
- **Channel capacity.** Highest upper bound on the rate of information that can be reliably transmitted/processed over a communications channel, and/or without causing bottlenecks.

- **Power consumption.** The amount of electricity consumed. This becomes especially important for systems with limited power sources, such as those deployed at the edge. (Power use (Watts), kW/h, etc.).
- **Compression ratio.** Data compression is subject to a space–time-complexity trade-off, which aims to reduce resource usage (data compression ratio, space savings, transmission savings, losses, etc.).
- **Environmental impact.** Measurements (such as power usage effectiveness (PUE), compute power efficiency (CPE), etc.) that are taken with the goals of decreasing waste, hazardous chemicals, and a computer’s ecological footprint are referred to as “green computing” measures.

5.1.3. Response Time

The response-time criteria evaluate the speed of a cyber situational awareness solution for being able to observe, deduce threats and/or risks and suggest the best suitable courses of action (CoAs). The response time takes into account the fact that some processes need a preceding observation period before the processing of the collected data, unlike measurements linked to performance and efficiency. As a result, the time taken to process the data adds to the delay in data collecting. The following illustrates the main measurable factors to be considered in the context of assessing the response time of processes of cyber-situational-awareness-related solutions.

- **Average response time.** The average response time (ART) per CSA operation is typically affected when slow inference actions are being conducted. It serves as a finer grain measurement, which usually require additional indicators to consider non-obvious problems in depth.
- **Peak response time.** The longest responses to all requests sent by the server are measured as the peak response time (PRT). This is a reliable indicator of how well CSA-related functionalities are performing.
- **Total raw response time.** The total raw response time (TRRT) is the sum of the response times at a certain time interval, which became very useful for assessing finer grain actions composed by several system procedures.
- **Response time percentile.** A percentile (or a centile) is a measure originated in statistics that indicates the range in which a given percentage of observations in a group of observations falls. For example, the 80th percentile of the response time is the value below which 80% of the conducted response time may be found. In the scope of threat/risks management and response to decision making, this may be an indicator of the proper processing of potential incidents based on prioritization.
- **Response time jitter.** Jitter in electronics and telecommunications is the divergence of a supposedly periodic signal from genuine periodicity, frequently in connection to a reference clock signal. Overall, jitter measures how much a specific action’s response time can vary.

5.1.4. Updatability and Upgradability

Updatability refers to the capability of modifying or adding functionalities, knowledge or functionalities to a system; being the upgradability a particular case where an original feature is fully replaced by a new one. In the context of cyber-situational-awareness acquisition, the term updatability refers to the capability to modify components, functionalities, configurations or data without disrupting CSA operations while minimizing the related maintenance costs. In analogy with the truth-functional propositional logic, updatability heavily lies in the feasibility of describing and enforcing rules of replacement, which shall fulfil commutativity, associativity, specificity and suitable updating costs. Beyond the state of the art, this methodology introduces the following novel updatability/upgradability indicators:

- **Commutativity.** Adapted to systems, services, knowledge, configuration, etc.; when commuting components, their results shall be independent of order. For example, it

shall be possible to digest new cyber threat intelligence (CTI) from a recent discovery source without their order (space-temporal) affecting the CSA risk-identification and management outcomes. Commutativity may be revealed at system tests where the divergence between different CTI organizational settings shall approach no differences. Regarding CSA capabilities, commutativity may be measured based on the similarity between organizational cases, potentially adopting similarity distances on the metrics adopted for assessing the capabilities of the compared cases separately (they can be evaluated in terms of accuracy, performances, etc.).

- **Associativity.** In the context of updating/upgrading cyber-situational-awareness-related capabilities, associativity refers to the degree to which how their internal functionalities as black-box views are not affected by the order or prioritization to be conducted. For example, a decision-support system may analyze concurrent if-then scenarios before suggesting the best suitable CoAs. Under an optimal associative condition, the order in which the scenarios are simulated and evaluated shall not affect the suggested decisions. Consequently, it is possible to state that this decision-support capability is more easy to update/upgrade when meeting the associativity property, since fewer conditions should be taken into account prior to conducting modifications. Similar to commutativity, associativity can measure the similarity between different execution cases, potentially adopting similarity distances on the metrics adopted for assessing the capabilities of the compared cases separately.
- **Discernibility.** Despite the proper satisfaction of the rest of the updatability/upgradability properties, a cyber-situational-awareness-acquisition solution can only be modified or updated when suitable factual and operational knowledge exists and may enhance its original capabilities. These “replacements” may be imported for external sources (e.g., COTS, OSSINT, etc.) and/or manufactured/developed based on the previous capabilities. In this context, discernibility refers to the existence and availability of such prior knowledge. For example, the lack of data for training/validation purposes typically entails a well-known problem for further add ons of machine learning. The CSA components mostly affected by this problem will present a lower level of discernibility than those that leverage data-rich environments and use data-driven algorithms.
- **Reversibility.** As considered in maths, reversibility refers to the existence of a functionality able to reverse the preliminarily achieved results regarding domain/range. Extending this definition for evaluating the updatability/upgradability of a cyber-situational-awareness system/sub-system, reversibility refers to the capability of achieving changes driven by updates/upgrades on the target of the evaluation. It is assumed that the degree of reversibility is maximized when updates/upgrades can completely change a functionality outcome. Similar to commutativity, reversibility can be assessed by measuring the similarity between different execution cases, potentially adopting similarity distances on the metrics adopted for assessing the capabilities of the compared cases separately.
- **Acquisition plan.** The evaluation of the acquisition plan for an update/upgrade is discomposed into two main features: total cost of acquisition (TCA) and acquisition model (AM). TCA entails estimations of the optimistic, pessimistic, and average cost, which is the total of the closing, research, accounting, commissions, legal fees, shipping, preparation, and installation costs of the purchase. On the other hand, AM defines the business model associated with the provisioning of upgrading/updating material, which may provide a greater/lower value based on the CSA tool capacitation context (ranging from free of economical and legal boundaries, up to direct sale of licenses, subscription models, or transaction-based charges).

5.1.5. Scalability

A cyber-situational-awareness acquisition system is said to be scalable if it can expand to meet rising computing demands and monitoring circumstances. A scalable system must be able to adjust to these changes without having a substantial negative influence on its

ability to process data, as this circumstance frequently directly affects this capacity. The ability to include new functionality or the simplicity of adapting to drastically increased workload rates are two ways that scalability can be measured. As a result, the factors to be considered while evaluating scalability are directly related to how the system is implemented as well as how it is integrated into the operational environment.

Many of the measures historically used for this purpose are compiled in [76]. However, in spite of their great variety, as indicated by Xiong et al. [77] currently, there is no one criterion for evaluating their applicability; rather, each proposal examines scalability from a certain angle and is concentrated on a particular use case. Scalability testing is a non-functional test methodology in which an application’s performance is measured in terms of its ability to scale up (vertical scalability) or scale down (horizontal scalability) the number of user requests or other such performance-measure attributes regarding the scalability target: performance, availability, maintenance, expenditures, etc.

With the purpose of assessing the scalability of cyber-situational-awareness-related systems and functionalities, and given the heterogeneity of features to be considered for scalability analytics, this methodology adopts an approach derived from queue theory research: an adaptation of Little’s law, the coefficient of efficiency and the zero-delay rate.

- **Outrun.** According to Little’s law [78], under steady-state conditions, the average number of items in a queuing system equals the average rate at which items arrive multiplied by the average time that an item spends in the system. Let us consider a queuing process where L is number of users, transactions, processes, etc. scheduled; W is the average waiting time; and λ is the average number of items arriving per unit of time, Little’s law describes the following relationship: $L = \lambda xw$. In the context of CIS, this expression is typically abstracted as a black-box view as follows: $X = \frac{N}{R}$; where X returns the outrun metric associated to the scalability throughput, N is the average number of operations conducted by the system/functionality target of evaluation; and R is the average operation duration.
- **Coefficient of efficiency.** The coefficient of efficiency p acts as a stochastic representation of the outrun metric, being detailed as follows: $p = \frac{\lambda}{m\mu}$; where λ is the average total of request per service, μ is the total capacity of the system/functionality for service, and m is the number of servers, resources, etc. which the target system is able to escalate with. Note that when $p > 1$ the evaluated capability is not able to property serve all the services so the request queue grows.
- **Zero delay rate.** Furthermore, the probability that a request has zero delay in a queue before receiving service is shown with $W(0)$. Equivalently, $1 - Q(0)$ is the probability that a customer has a nonzero delay. The formula that gives the probability that an arriving customer is delayed in the queue (i.e., has positive, non-zero wait in the queue), is a function of the parameters c and r , $r = \frac{\lambda}{\mu}$ called the Erlang-C formula [79].

$$C(m, r) = 1 - q(0) = \frac{\frac{r^m}{m!(1-p)}}{\frac{r^m}{m!(1-p)} + \sum \frac{r^n}{n!}} \tag{2}$$

5.1.6. Robustness against Adversarial Tactics

Targeted attacks against these systems are now more lucrative due to the recent rise in popularity of new disruptive technologies and the rapid development of a digital society. This also facilitated the development of a number of evasion techniques capable of nullifying even the most successful intrusion-detection and risk-management techniques [80]. The necessity to create effective detection methods that can withstand denial-of-service attacks and detect malicious actions concealed by imitation observations of appropriate behaviors is a direct result of these new threats [81–83]. However, despite their significant and increasing evolution, there are no standardized procedures or metrics for assessing the resistance of systems, services, functionalities, etc. against these threats. As a response, this methodology proposes to follow the same procedure adopted by most of the state-of-the-art

research publications, which rely on conducting comparisons between the effectiveness achieved by conventional scenarios against the results observed when repeating their execution modified by evasion/adversarial tactics. With this purpose, this methodology introduces the effectiveness loss (EL) metrics, which is calculated as follows:

$$EL = [Effectiveness\ in\ normal\ case] - \frac{Effectiveness\ under\ evasion}{adversarial\ tactics} \quad (3)$$

where the effectiveness of a capability shall be properly measured according to its dimension (e.g., accuracy, performance, scalability, etc.) and the type of operation it entails (i.e., in the grounds of the most suitable associated metrics). Accordingly, the greater the effectiveness loss, the lesser the capability is hardened against adversarial tactics.

5.2. Cyber-Threat Mitigation on the Mission Plane

There are not consolidated frameworks for evaluating the implementation of mission-centered cyber-risk-management-related capabilities. In this context, the proposed framework suggests the adoption of conventional solutions for well-known methodologies, such as Magerit [84] or CVSS [85], as the evaluation framework developed by the ENISA working group (WG3) on risk assessment and risk management (RA/RM) [86]. This paper abstracted this methodology to the particularities of the military operations on joint domains aligned with the ISO 31,000 method, as illustrated in Figure 6.

- **P1. Definition of external environment.** Statement of all those capabilities for the mission that have a direct dependence on the outside world, i.e., all those capabilities whose functioning may be altered by external elements to the mission.
- **P2. Definition of internal environment.** Statement of all mission capabilities whose behavior may be altered by an internal failure. Internal faults are considered as all those faults that do not have an external origin and can affect to the rest of the elements of the mission.
- **P3. Generation of risk management context.** Statement of how risks and threats (external or internal) that have a direct impact on some mission capability (in the first instance) and/or on some of the elements of the mission will be managed.
- **P4. Formulation of impact limit criteria.** Statement of how risks are spread hierarchically and vertically/horizontally propagated. Thus, it is necessary to define, firstly, the propagation from the CIS capability plane to the operational task plane and, secondly, from the task plane to the mission plane. For this purpose, solutions such as Bayesian Networks are easily configurable, versatile and extensible.
- **P5. Identification of risks.** During this stage of the methodology, it is necessary to break down the mission into different tasks, and identify the high-level risks that may affect to the mission and how these risks are distributed among the different tasks.
- **P6. Analysis of relevant risks.** This task lowers the abstraction level in the methodology process and seeks to understand how CIS capabilities affect each of the elements involved in each of the tasks of the mission plane. To this end, it is proposed to make radar diagrams to identify the impact that different variables or dimensions have on each of the elements of the mission. The dimensions to be evaluated in order to quantify risks are the different dimensions of the vulnerability and threat-assessment methodology (CVSS v3.1) [85].
- **P7. Evaluation of risks.** Once the impact value of the different threats (P6) has been quantified, it is necessary to calculate the mission-level risks. To do this, it will be necessary not only to calculate the risk of each element but also its implication in the different tasks of the mission and its final impact on the mission goals, being calculated as the target of evaluation tools suggests.
- **P8. Identification of options.** Identification of the spread of risk between mission tasks as severity of aggregated risk at the mission level.
- **P9. Development of action plan.** Definition of action or contingency plans, i.e., business continuity and recovery plans (BCP), for each of the previously defined risks.

These action plans should be defined for changes from lower levels to the CIS level, so that in the future the system can infer dependencies between different tasks.

- **P10. Approval of action plan.** Approval of the contingency action plan for the current active risk in order to minimize their impact on the mission.
- **P11. Implementation of action plan.** Implementation of the action plan approved in the previous stage.
- **P12. Identification of residual risks.** New analysis of the current situation to re-identify risks. It is especially important to analyze the impact between tasks and residual risks.
- **P13. Risk acceptance.** Definition of tolerance thresholds to discriminate whether or not to apply corrective measures (business plan actions) and whether or not to abort the mission.
- **P14. Risk monitoring and reporting.** This stage is related with the system of visualization and control of the state of the risks at the mission [87]. It is in charge of generating different reports of the current situation to facilitate the decision-making process. In addition to this, the system could suggest a set of the decisions to be taken and the possible impact they will have on the correct development of the mission.
- **P15. Risk communication, awareness and consulting.** The last process of the methodology is related with the communication, awareness and consulting tasks. Hence, this module must interact with the different elements of the full system, intercommunicate, and propagate the risk information.

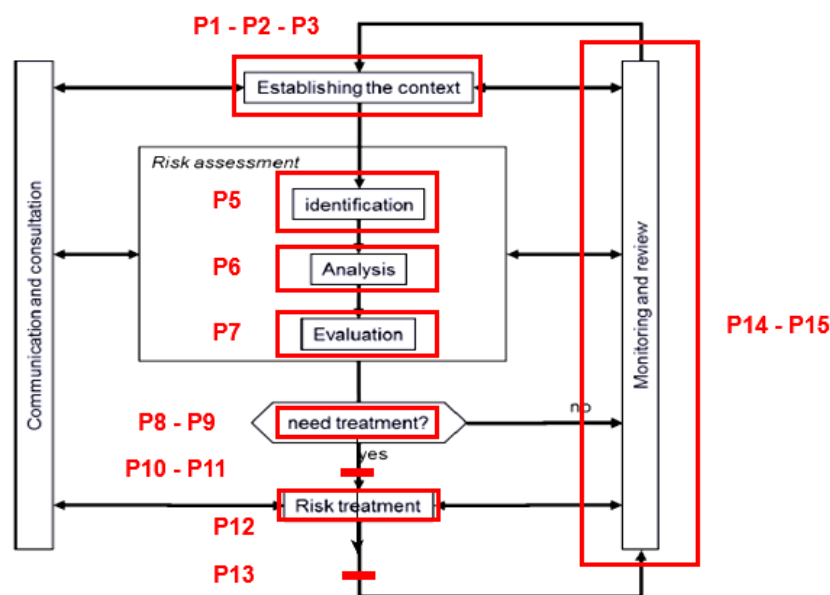


Figure 6. Integration of ENISA’s guidelines [86] to ISO31000.

5.3. Support to Decision-Making

Designing an appropriate framework [88] to ensure proper defence against various potential attacks is not a simple task. There is a large research stream devoted exclusively to this area, wherein evaluations of decision-making capabilities typically rely on the basis of the operational research. Part of the difficulty associated with their design is that they require a very precise understanding of the global context in which CSA acquisition tools operate, ranging from sensors, enablers up to cyber command and control (C2) systems. Overall, they are in-charge of protecting CIS assets by facilitating the selection and planning of the most appropriate cyber CoAs. Accordingly, every decision-making stage must be carried out based on a previous set of acquired cyber intelligence (for example, IDS reports) and enforced driven by sets of possible countermeasures to minimize the effect of the threats propagated to the mission plane.

The current literature distinguishes two different trends depending on how they pose the problem. The first approaches decision making driven by dependency graphs that model the entire test scenarios [89]. Thus, potential risks are segmented into different stages and the environment is modelled in order to seek the least impact at all times. The second approach is based on treating the decision-making process as a multidimensional optimization problem, in which each dimension corresponds to a different criterion. This approach is more complex to model and assess, since all the performance indicators to be considered shall be previously established and evaluated [90]. Due to this, the instantiation of the CSA evaluation framework shall suggest exploring hybrid solutions similar to [90], but adapted to the mission-centric context, where for each decision dimension, a weight is defined in order to allow the balancing of the multidimensional criteria. These weights are normalized, and their definition must be conducted during the planning stage of the missions, specifically, during the definition of possible CoAs.

6. Cyber Situational Awareness Application Concepts

The following describes the proposal verification and validation activities related with the application of the CSA-related cyber defence tools, covering both user acceptance and their effectiveness when supporting the human adequacy for the acquisition of CSA.

6.1. Capability of Facilitating CSA Acquisition

In order to evaluate the degree of knowledge acquired during the development of a mission, it is necessary to know beforehand which external and internal agents can influence the assimilation and processing of the information. Among these agents, all those that affect in a direct or indirect way the psychological component of people stand out. Generally, anxiety and stress are two of the factors with which we will have to struggle most often. Through practice and training, it is possible to minimize their impact on decision making. However, due to the high degree of subjectivity that this type of approach has, it is necessary to carry out an evaluation methodology segmented into three different stages: baseline, test performance, and final assessment. In the baseline, all the measurement systems used in the solution must be calibrated in order to build a more reliable and contrasted information. To this end, the following practices can be followed: (1) to present the tasks to be performed; (2) to play quiet music to bring the subject to the baseline more quickly; and (3) to conduct initial questionnaires to assess in a subjective way the initial perceptions of the individual (e.g., SAGAT pre-mission assessment [59]). The performance of the test corresponds to the second stage of the methodology, which relies on simulations of situations similar to those the individual will encounter during military operations. During the simulations all the raw data (for later study) and questionnaires will be collected and validated. The process of returning the individual to the basal levels is studied during the last stage. Among others, it is recommended to ask a final survey on the process followed in order to measure and assess the subjective perception of the operational environment and the accuracy of the acquired operational picture (e.g., SART [59]).

6.2. User Acceptance

The willingness of a user group (experts or decision makers) to use new technologies for the tasks they are intended to serve is known as user acceptance. As a result, this idea is not being applied to cases where users promise to utilise a technology without presenting any evidence of use or to the use of a technology for reasons that the inventors did not intend. The goal of the evaluation is to confirm that the CSA tools are appropriate for the end user. The desire of the concerned stakeholders to adopt and utilise the generated tools in their daily activities will also be taken into consideration when evaluating user acceptance, in addition to the project's scope and duration. The Questionnaire for User Interaction Satisfaction (QUIS) approach has been modified in the proposed methodology for suiting cyber situational awareness [91]. When it comes to comprehending situational awareness as a human "mental state," QUIS was developed as a measurement instrument

to gauge a computer user's subjective satisfaction with the human-computer interface. Accordingly, the following performance indicators for assessing the user acceptance level:

- **Usability.** Degree to which CSA tools can be used by end-users to achieve their tasks with effectiveness, efficiency, and satisfaction in a quantified context of use.
- **Ease of Use.** Learnability of the CSA tools, and the degree of intuitively for end users.
- **Terminology and System Information.** Degree to which the system notification, description and presentations are coherent with the CSA operational context.
- **Functionality.** The range and effectiveness of operations that can be conducted by the provided CSA.
- **Satisfaction.** Degree to which a CSA enabler fulfills the end-user expectations

The acceptance testing process begins with the execution of individual tests (known as test cases), on which users will be questioned, bearing in mind that: (1) test cases are executed, using predetermined data; (2) actual results are recorded; (3) actual and expected results are compared; and (4) test results are determined. The questionnaires ask questions to end users to estimate the value associated to each performance indicator, which shall be answered ranging from: 1—Strongly agree to 5—Strongly disagree. Questions include “Using the CSA capabilities would improve my job performance?”, “Learning to operate the CSA enablers would be easy for me”, or “Do the evaluated capabilities bring a clear picture of the cyberspace?”.

6.3. Acceptance Questionnaire

The questionnaire in Table 2 proposes some questions for an end user to estimate the value associated to each performance indicator, which shall be answered for the stated statements using a scale that goes from: 1—Strongly agree to 5—Strongly disagree. Note that these questions may serve just as guide that could be modified, extended or adapted to the system development of validation stage (e.g., alpha testing, beta testing, commercial acceptance, etc.). The score associated to each performance indicator typically is calculated as the mean, mode, median, and/or standard deviation of the rates collected, which will be analyzed assuming all the interviewed audience, or specific groups.

Figure 7 summarizes an activity diagram of the evaluation workflow. Accordingly, three lines of evaluation actions are discerned: testing, operations and applications. As illustrated, the actions concerning the testing concept are suggested to be executed according to the following sequence: unity tests, integrity tests, security tests and reliability tests. However, it will be up to the quality assurance (QA) and analyst teams to make any variations they deem appropriate based on the scope of testing or the resources available to them. This line is suggested to be enforced before assessing the operational effectiveness of the proposal, so the required tests are usually integrated independently, regarding the operational context (but assuming friction conditions of the operational concept on which the cyber defence tools operate). Due to this feature, they could be paralleled with any other line of action. However, any defects found in tests could invalidate the results at the operational or acceptance level, so it is therefore very important to have a minimum of compliance in testing before moving on to more complex assessment stages.

The operational concept is evaluated once the evaluation loop is triggered, including cross-component validations, and the analysis of the effectiveness of the solution discovery, risk assessment and risks management capabilities. The evaluation loop is triggered once per phase on the cyber-kill chain of each attack scenario. At the end of each phase the application concepts will be evaluated, including the capability for achieving cyber situational awareness and user acceptance. All the observed results will be properly collected and stored for supporting further modifications, integrations or deployments at different operational contexts. Operational assessments study the specific capabilities offered by each cyber defence tool, so the work of experts in defining them, integrating them and establishing compliance criteria will be essential to their effectiveness.

Table 2. Example of User Acceptance questionnaire.

		Rate				
		1	2	3	4	5
Usability						
U1	Using the CSA tool would enable me to accomplish tasks more quickly and effectively tasks					
U2	Using the CSA tool would improve my job performance					
U3	Using the CSA tool would make it easier to do my job					
U4	I would find the CSA tool useful in my job					
Ease of use						
E1	Learning to operate the CSA tool would be easy for me					
E2	It would be easy to get the CSA tool to assist in the fulfilment of tasks					
E3	My interaction with the CSA tool would be clear and understandable					
E4	I would find the CSA tool to be flexible to interact with					
E5	It would be easy for me to become skillful at using the CSA tool					
E6	I would find the CSA tool easy to use					
Terminology and System Information						
T1	I find the use of terms and concept of operation throughout the system adequate					
T2	I find the terminology related to mission, tasks, incidents and courses of actions adequate					
T3	The position of messages on screen is intuitive					
T4	The risk levels and their potential propagation are displayed comprehensively and according to their criticality					
T5	The CSA tool informs about its internal processes non-intrusively and as requested					
T6	Alerts are property emphasized					
Functionality						
F1	I found the various functions in the CSA tool well integrated					
F2	The CSA tool brings a clear picture of the cyberspace					
F3	The CSA tool brings a clear picture of the relationship between cyberspace and planned/ongoing missions					
F4	The CSA tool brings comprehensively and effective support to courses of action identification, selection and planning					
F5	The integration of the CSA tool with external data sources (SOC, NOC, CTI, Mission planners) is properly operative					
F6	The analytical capabilities integrated (simulation, diagnosis, prediction, etc.) are effective					
F7	The CSA tool is self-protected and implements a consisted audition system					
Satisfaction						
S1	I am satisfied with the CSA tool					
S2	I would like to use the CSA tool frequently					
S3	I would recommend the CSA tool to my team					
S4	The CSA tool works the way I want it to work					
S5	I would need the CSA tool for my daily tasks					
S6	I am very confident with the use of the CSA tool					

Finally, user acceptance will be assessed, which is expected to be dependent on to what extent the solutions to be evaluated are able to ease the acquisition of cyber situational awareness for which beyond the accuracy on the perceived operational pictures, questions such as usability, ease of use, satisfaction, terminology, functionality, etc. shall be measured. However, regarding this dependence, the authors do not suggest paragraphing the questionnaires, based on the proven effectiveness of the tool, in such a way that the outcome of acceptance indicators is less biased by operational results. I.e., for example, an

ad-hoc reformulation of the questionnaires based on a bad operational experience (probably with the intention to dig into what happened) most likely will lead to directing the user perception of the tools to be evaluated towards the specific aspects that led to the bad operation, which may be disconnected from topics to be considered in order to accept the solution in a different operational condition. Acceptance will be assessed by direct consultation to the cyber defence tools operators.

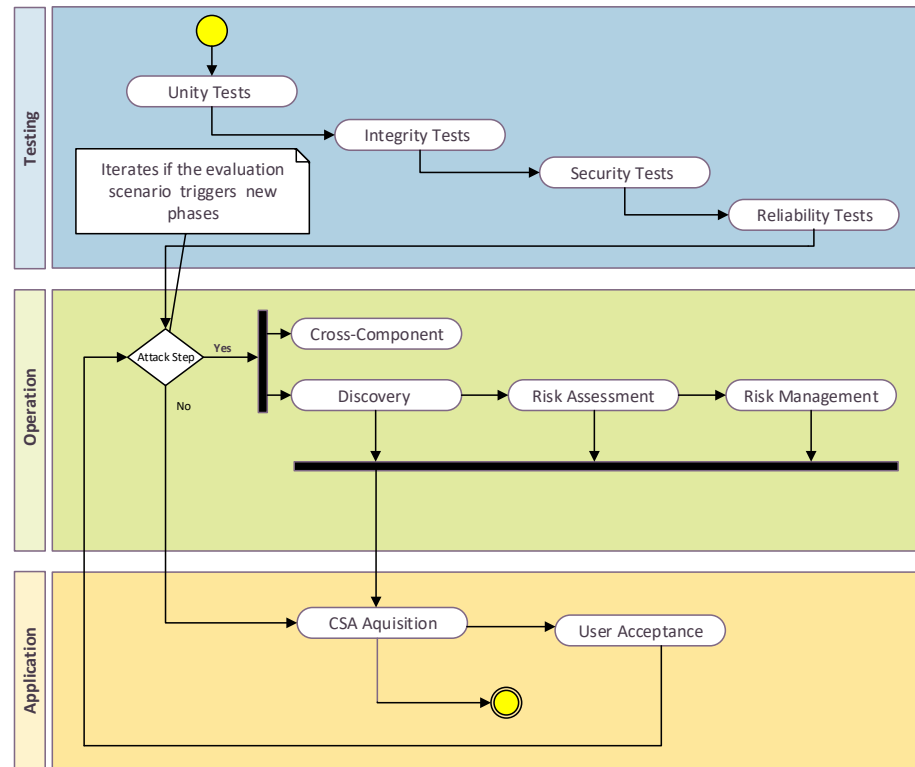


Figure 7. Action diagram of the evaluation workflow.

7. Conclusions

The presented research has delved into raising difficulties, challenges and gaps related with the evaluation of CSA acquisition tools. The assessment on capabilities for related dual-use solutions have been reviewed, concluding that the state-of-the-art lacks consolidated mission-centric CSA validation enablers. With the purpose of contributing to their development, a novel verification and validation framework to assist a proper evaluation has been introduced, which proposed three core assessment concepts: software tests, operations and applications. The first concept covers the proper technical implementation of the capabilities, the second concept describes the core functionalities for supporting CSA acquisition (perception of the operational environment, assessment of vertically propagated threats from cyberspace to the mission plane, and support to decision-making), and the third concept studies the applicability of the solutions in terms of users' acceptance and the quality of their acquired operational picture. The proposed method has been presented as a general-purpose mission-centric solution applicable to heterogeneous cyber defence tools, being open to expansion, modification and any other change that the singularities of a particular CSA enabler and its end-user operators require.

Although efforts have been made to cover all the perceived essential aspects, it is expected that further enhancements and upgrades will tentatively make it a better fit to the particularities of certain cross-cutting operational domains or functionalities, thus projecting its applicability in the short, mid and long term. The authors want to highlight that this paper brings together extensive research and synthesis work, in some cases with very few (or non-existent) precedents. Since cyber defence is an emerging research field still

with many challenges and technological/analytical gaps, it is expected that the presented research outcomes establish the grounds for future related works, as well as incentivising further research actions.

Author Contributions: All authors contributed to Conceptualization, Investigation, Design, Writing, Reviewing and Editing: S.L.S., D.S.R.-B., R.D.M., R.P.d.R., F.T. and J.M.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Disclaimer: The contents reported in the paper reflect the opinions of the authors and do not necessarily reflect the opinions of the respective agencies, institutions or companies.

References

- Adam, E. Fighter Cockpits of the Future. In Proceedings of the 12th IEEE Digital Avionics Systems Conference, Fort Worth, TX, USA, 25–28 October 1993; pp. 318–323.
- Dahal, N.; Abuomar, O.; King, R.; Madani, V. Event Stream Processing for Improved Situational Awareness in the Smart Grid. *Expert Syst. Appl.* **2015**, *42*, 6853–6863. [[CrossRef](#)]
- Endsley, M.; Selcon, S.; Hardiman, T.; Croft, D. A Comparative Analysis of Sagat and Sart for Evaluations of Situation Awareness. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Chicago, IL, USA, 5–9 October 1998; pp. 82–86.
- Bigelow, B. What are Military Cyberspace Operations Other Than War? In Proceedings of the 11th IEEE International Conference on Cyber Conflict (CyCon), Tallinn, Estonia, 28–31 May 2019; pp. 318–323.
- Lif, P.; Granasen, M.; Sommestad, T. Development and validation of technique to measure cyber situation awareness. In Proceedings of the International Conference On Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA), London, UK, 19–20 June 2017; pp. 1–8.
- Maestre Vidal, J.; Orozco, A.; Villalba, L. Adaptive artificial immune networks for mitigating DoS flooding attacks. *Swarm Evol. Comput.* **2018**, *38*, 94–108. [[CrossRef](#)]
- Maestre Vidal, J.; Sotelo Monge, M. A novel Self-Organizing Network solution towards Crypto-ransomware Mitigation. In Proceedings of the 13th International Conference on Availability, Reliability and Security (ARES), Hamburg, Germany, 27–30 August 2018.
- Sandoval Rodriguez-Bermejo, D.; Daton Medenou, R.; Ramis Pasqual de Riquelme, G.; Maestre Vidal, J.; Torelli, F.; Llopis Sánchez, S. Evaluation methodology for mission-centric cyber situational awareness capabilities. In Proceedings of the 15th International Conference on Availability, Reliability and Security (ARES), Virtual Event, Ireland, 25–28 August 2020; pp. 1–9.
- Endsley, M. Towards a theory of situational awareness in dynamic systems. *Hum. Factors* **1995**, *37*, 32–64. [[CrossRef](#)]
- Chatzimichailidou, M.; Stanton, N.; Dokas, I. The Concept of Risk Situation Awareness Provision: Towards a New Approach for Assessing the DSA about the Threats and Vulnerabilities of Complex Socio-Technical Systems. *Saf. Sci.* **2015**, *79*, 126–138. [[CrossRef](#)]
- Barona López, L.; Valdivieso Caraguay, A.; Maestre Vidal, J.; Sotelo Monge, M. Towards Incidence Management in 5G Based on Situational Awareness. *Future Internet* **2017**, *9*, 3. [[CrossRef](#)]
- Lenders, V.; Tnner, A.; Blarer, A. Gaining an Edge in Cyberspace with Advanced Situational Awareness. *IEEE Secur. Priv.* **2015**, *13*, 65–74. [[CrossRef](#)]
- Franke, U.; Brynnielsson, J. Cyber Situational Awareness—A Systematic Review of the Literature. *Comput. Secur.* **2014**, *46*, 18–31. [[CrossRef](#)]
- Saadou, A.; Chenji, H. Optimizing Situational Awareness in Disaster Response Networks. *IEEE Access* **2018**, *6*, 24625–24638. [[CrossRef](#)]
- Webb, J.; Ahmad, A.; Maynard, S.; Shank, G. A Situation Awareness Model for Information Security Risk Management. *Comput. Secur.* **2014**, *44*, 1–15. [[CrossRef](#)]
- Silva, J.A.H. and Lopez, L.; Caraguay, A.; Hernández-Álvarez, M. A Survey on Situational Awareness of Ransomware Attacks—Detection and Prevention Parameters. *Remote Sens.* **2019**, *11*, 1168. [[CrossRef](#)]
- Ioannou, G.; Louvieris, P.; Clewley, N. A Markov Multi-Phase Transferable Belief Model for Cyber Situational Awareness. *IEEE Access* **2019**, *7*, 39305–39320. [[CrossRef](#)]
- Elbez, H.; Keller, H.; Hagenmeyer, V. A New Classification of Attacks against the Cyber-Physical Security of Smart Grids. In Proceedings of the 13th International Conference on Availability, Reliability and Security, Hamburg, Germany, 27–30 August 2018; p. 63.

19. Bolbont, V.; Theotokatos, G.; Bujorianu, K.; Boulougouris, E.; Vassalos, D. Vulnerabilities and Safety Assurance Methods in Cyber-Physical Systems: A Comprehensive Review. *Reliab. Eng. Syst. Saf.* **2019**, *182*, 179–193. [[CrossRef](#)]
20. Demertzis, K.; Tziritas, N.; Kikiras, P.; Llopis Sanchez, S.; Iliadis, L. The Next Generation Cognitive Security Operations Center: Network Flow Forensics Using Cybersecurity Intelligence. *Big Data Cogn. Comput.* **2018**, *2*, 35. [[CrossRef](#)]
21. Kriaa, S.; Petre-Cambacedes, L.; Bouissou, M.; Halgand, Y. A Survey of Approaches Combining Safety and Security for Industrial Control Systems. *Reliab. Eng. Syst. Saf.* **2015**, *139*, 156–176. [[CrossRef](#)]
22. Fantini, P.; Pinzone, M.; Taisch, M. Placing the Operator at the Centre of Industry 4.0 Design: Modelling and Assessing Human Activities within Cyber-Physical Systems. *Comput. Ind. Eng.* **2020**, *139*, 105058. [[CrossRef](#)]
23. Gharib, M.; Lollini, P.; Ceccarelli, A.; Bondavalli, A. Dealing with Functional Safety Requirements for Automotive Systems: A Cyber-Physical-Social Approach. In Proceedings of the 12th International Conference on Critical Information Infrastructures Security, Lucca, Italy, 8–13 October 2017; pp. 194–206.
24. Sotelo Monge, M.A.; Maestre Vidal, J.; Martínez Pérez, G. Detection of economic denial of sustainability (EDoS) threats in self-organizing networks. *Comput. Commun.* **2019**, *145*, 284–308. [[CrossRef](#)]
25. Zeng, J.; Yang, L.; Lin, M.; Ning, H.; Ma, J. A Survey: Cyber-Physical-Social Systems and their System-Level Design Methodology. *Future Gener. Comput. Syst.* **2020**, *105*, 1028–1042 [[CrossRef](#)]
26. Wang, P.; Yang, L.; Li, J.; Hu, S. Data Fusion in Cyber-Physical-Social Systems: State-of-the-Art and Perspectives. *Inf. Fusion* **2016**, *51*, 42–57. [[CrossRef](#)]
27. Llopis Sanchez, S.; Mazzolin, R.; Kechaoglou, I.; Wiemer, D.; Mees, W.; Muylaert, J. Cybersecurity Space Operation Center: Countering Cyber Threats in the Space Domain. In *Handbook of Space Security*; Springer: Cham, Switzerland, 2019. [[CrossRef](#)]
28. Fortson, L.W. Towards the Development of a Defensive Cyber Damage and Mission Impact Methodology. Master's Thesis, Air Force Institute of Technology, Kaduna, Nigeria, 2007.
29. Demertzis, K.; Tziritas, N.; Kikiras, P.; Llopis Sanchez, S.; Iliadis, L. The Next Generation Cognitive Security Operations Center: Adaptive Analytic Lambda Architecture for Efficient Defense against Adversarial Attacks. *Big Data Cogn. Comput.* **2019**, *3*, 6. [[CrossRef](#)]
30. Price, P.; Leyba, N.; Gondreey, M.; Staples, Z.; Parker, T. Asset criticality in mission reconfigurable cyber systems and its contribution to key cyber terrain. In Proceedings of the 50th International Conference on Systems Sciences (HICSS 2017), Waikoloa Village, HI, USA, 4–7 January 2017; pp. 446–456.
31. Schulz, A.; Kotson, M.; Zipkin, J. *Cyber Network Mission Dependencies*; Technical Report 1189; Massachusetts Institute of Technology, Lincoln Laboratory: Lexington, MA, USA, 2015.
32. Cheng, M-Wang, B.; Zhao, S.; Zhai, Z.; Zhu, D.; Chen, J. Situation-Aware Dynamic Service Coordination in an IoT Environment. *IEEE/ACM Trans. Netw.* **2017**, *25*, 2082–2095. [[CrossRef](#)]
33. Cohen, G.; Afshar, S.; Morreale, B.; Bessell, T.; Wabnitz, A.; Rutten, M.; van Schaik, A. Event-based Sensing for Space Situational Awareness. *J. Astronaut. Sci.* **2019**, *66*, 125–141. [[CrossRef](#)]
34. Layton, P. Fifth-generation air warfare. *Aust. Def. Force J.* **2018**, *204*, 23–32.
35. de Barros Barreto, A.; Costa, P.; Yano, E. Using a semantic approach to cyber impact assessment. In Proceedings of the 8th Conference on Semantic Technologies for Intelligence, Defense, and Security (STIDS 2013), Fairfax, VA, USA, 13–14 November 2013; pp. 101–108.
36. D'Amico, A.; Buchanan, L.; Goodall, J.; Walczak, P. *Mission Impact of Cyber Events: Scenarios and Ontology to Express the Relationships between Cyber Assets, Missions, and Users*; Tech. Rep. OMB No. 0704-0188; AFRL/RIEF, US Defence Technical Information Center; Fort Belvoir, VA, USA, 2009.
37. Endsley, M. Situational awareness misconceptions and misunderstanding. *J. Cogn. Eng. Decis. Mak.* **2016**, *9*, 4–32. [[CrossRef](#)]
38. Brynielsson, J.; Franke, U.; Varga, S. *Cyber Situational Awareness Testing. Combatting Cybercrime and Cyberterrorism*; Springer: Cham, Switzerland, 2016; pp. 209–233.
39. Stevens, S. Measurement, Statistics, and the Schemapiric View. *Science* **1968**, *161*, 849–856. [[CrossRef](#)]
40. Parasuraman, R.; Sheridan, T.; Wickens, C. Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs. *J. Cogn. Eng. Decis. Mak.* **2018**, *2*, 140–160. [[CrossRef](#)]
41. Dekker, S.; Hummerdal, D.; Smith, K. Situation awareness: Some remaining questions. *Theor. Issues Ergon. Sci.* **2008**, *11*, 131–135. [[CrossRef](#)]
42. Salmon, M.; Stanton, N.; Walker, G.; Baber, C.; Jenkins, D.; McMaster, R.; M.S., Y. What really is going on? Review of situation awareness models for individuals and teams. *Theor. Issues Ergon. Sci.* **2008**, *9*, 297–323. [[CrossRef](#)]
43. Endsley, M. A Survey of Situation Awareness Requirements in Air-to-Air Combat Fighters. *Int. J. Aviat. Psychol.* **1993**, *3*, 157–168. [[CrossRef](#)]
44. Endsley, M. Situation awareness global assessment technique (SAGAT). In Proceedings of the IEEE National Aerospace and Electronics Conference, Dayton, OH, USA, 23–27 May 1988; pp. 937–942.
45. Salmon, P.; Stanton, N.; Jenkins, D. How Do We Know What They Know? Situation Awareness Measurement Methods Review. In *Distributed Situation Awareness*; CRC Press: Boca Raton, FL, USA, 2017; pp. 55–76.
46. Miles, J.; Strybel, T. Measuring Situation Awareness of Student Air Traffic Controllers with Online Probe Queries: Are We Asking the Right Questions? *Int. J. Hum. Comput. Interact.* **2017**, *33*, 55–65. [[CrossRef](#)]

47. Matthews, M.; Beal, S. *Assessing Situation Awareness in Field Training Exercises*; Research Report 1795; West Point US Military Academy: West Point, NY, USA 2002.
48. Tounsi, W.; Rais, H. A Survey on Technical Threat Intelligence in the Age of Sophisticated Cyber Attacks. *Comput. Secur.* **2018**, *72*, 212–233. [[CrossRef](#)]
49. Trautsch, F.; Herbold, S.; Grabowski, J. Are unit and integration test definitions still valid for modern Java projects? An empirical study on open-source projects. *J. Syst. Softw.* **2010**, *159*, 110421. [[CrossRef](#)]
50. Endsley, M. Measurement of situation awareness in dynamic systems. *Hum. Factors J. Hum. Factors Ergon. Soc.* **1995**, *37*, 65–84. [[CrossRef](#)]
51. Buczak, A.; Guven, E. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. *IEEE Commun. Surv. Tutor.* **2015**, *18*, 1153–1176. [[CrossRef](#)]
52. Gutzwiller, R.; Hunt, S.; Lange, D. A task analysis toward characterizing cyber-cognitive situation awareness (CCSA) in cyber defense analysts. In Proceedings of the IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), San Diego, CA, USA, 21–25 March 2016.
53. Mahoney, S.; Roth, E.; Steinke, K.; Pfautz, J.; Wu, C.; Farry, M. A Cognitive Task Analysis for Cyber Situational Awareness. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, San Francisco, CA, USA, 27 September–1 October 2010; Volume 54, pp. 279–283.
54. Ben-Asher, N.; Gonzalez, C. Effects of cyber security knowledge on attack detection. *Comput. Hum. Behav.* **2015**, *48*, 51–61. [[CrossRef](#)]
55. Mancuso, V.; Christensen, J.; Cowley, J.; Finomore, V.; Gonzalez, C.; Knott, B. Human Factors in Cyber Warfare II. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Chicago, IL, USA, 27–31 October 2014; Volume 58, pp. 415–418.
56. Tenney, Y.; Pew, R. Situation Awareness Catches On: What? So What? Now What? *Hum. Factors Ergon.* **2006**, *2*, 1–34. [[CrossRef](#)]
57. Malviya, A.; Fink, G.; Segó, L.; Endicott-Popovsky, B. Situational Awareness as a Measure of Performance in Cyber Security Collaborative Work. In Proceedings of the 8th International Conference on Information Technology: New Generations, Las Vegas, NV, USA, 11–13 April 2011; pp. 937–942.
58. Giacoe, N. Measuring the Effectiveness of Visual Analytics and Data Fusion Techniques on Situation Awareness in Cyber-Security. Ph.D. Thesis, The Pennsylvania State University, State College, PA, USA, 2012.
59. Evangelopoulou, M.; Johnson, C. Attack Visualisation for Cyber-Security Situation Awareness. In Proceedings of the 9th IET International Conference on System Safety and Cyber Security, Manchester, UK, 15–16 October 2014; pp. 937–942.
60. Fink, G.; Best, D.; Manz, D.; Popovsky, V.; Endicott-Popovsky, B. Gamification for Measuring Cyber Security Situational Awareness. In Proceedings of the International Conference on Augmented Cognition, Las Vegas, NV, USA, 21–26 July 2013.
61. Shiravi, H.; Shiravi, A.; Ghorbani, A. A Survey of Visualization Systems for Network Security. *IEEE Trans. Vis. Comput. Graph.* **2012**, *18*, 1313–1329. [[CrossRef](#)]
62. Dressler, J.; Bowen, C.; Moody, W.; Koepke, J. Operational data classes for establishing situational awareness in cyberspace. In Proceedings of the 6th International Conference On Cyber Conflict (CyCon 2014), Tallinn, Estonia, 3–6 June 2014; pp. 175–186.
63. Katasonov, A.; Sakkinen, M. Requirements quality control: A unifying framework. *Requir. Eng.* **2006**, *11*, 42–57. [[CrossRef](#)]
64. Zimek, A.; Vreeken, J. The blind men and the elephant: On meeting the problem of multiple truths in data from clustering and pattern mining perspectives. *Mach. Learn.* **2015**, *98*, 121–155. [[CrossRef](#)]
65. Jackson, M. *Software Requirements & Specifications: A Lexicon of Practice, Principles and Prejudices*; CM Press/Addison-Wesley Publishing Co.: New York, NY, USA, 1995.
66. M., E.; Shoval, S.; Joiner, K. System Test Architecture Evaluation: A Probabilistic Modeling Approach. *IEEE Syst. J.* **2019**, *13*, 3651–3662. [[CrossRef](#)]
67. Felderer, M.; Buchler, M.; Johns, M.; Brucker, A.; Breu, R.; Pretschner, A. Chapter One—Security Testing: A Survey. *Adv. Comput.* **2016**, *101*, 1–51.
68. OWASP. Open Web Application Security Project. 2020. Available online: <https://owasp.org/> (accessed on 8 June 2022).
69. ANSI/IEEE. *Standard Glossary of Software Engineering Terminology*; STD-729-1991; ANSI/IEEE: New York, NY, USA, 1991.
70. Bhuyan, M.; Bhattacharyya, D.; Kalita, J. Network Anomaly Detection: Methods, Systems and Tools. *IEEE Commun. Surv. Tutor.* **2014**, *16*, 303–336. [[CrossRef](#)]
71. van Schaik, P.; Renaud, K.; Wilson, C.; Jansen, J.; Onibokun, J. Risk as affect: The affect heuristic in cybersecurity. *Comput. Secur.* **2020**, *90*, 101651. [[CrossRef](#)]
72. Maestre Vidal, J.; Castro, J.; Orozco, A.; Villalba, L. Evolutions of evasion techniques against network intrusion detection systems. In Proceedings of the 6th International Conference on Information Technology, Bangkok, Thailand, 12–13 December 2013.
73. Maestre Vidal, J.; Sotelo Monge, M. Obfuscation of Malicious Behaviors for Thwarting Masquerade Detection Systems Based on Locality Features. *Sensors* **2020**, *20*, 2084. [[CrossRef](#)] [[PubMed](#)]
74. Huancayo Ramos, K.; Sotelo Monge, M.; Maestre Vidal, J. Benchmark-Based Reference Model for Evaluating Botnet Detection Tools Driven by Traffic-Flow Analytics. *Sensors* **2020**, *20*, 4501. [[CrossRef](#)] [[PubMed](#)]
75. Maestre Vidal, J.; Sotelo Monge, M.; Villalba, L. A novel pattern recognition system for detecting Android malware by analyzing suspicious boot sequences. *Knowl.-Based Syst.* **2018**, *150*, 198–217. [[CrossRef](#)]
76. Jogalekar, P.; Woodside, M. Evaluating the scalability of distributed systems. *IEEE Trans. Parallel Distrib. Syst.* **2000**, *11*, 589–603. [[CrossRef](#)]

77. Xiong, H.; Zeng, G.; Zeng, Y.; Wang, W.; Wu, C. A novel scalability metric about iso-area of performance for parallel computing. *J. Supercomput.* **2014**, *68*, 652–671. [[CrossRef](#)]
78. Salmanian, Z.; Izadkhah, H.; Isazdeh, A. Optimizing web server RAM performance using birth–death process queuing system: scalable memory issue. *J. Supercomput.* **2017**, *73*, 5221–5238. [[CrossRef](#)]
79. Gross, D.; Shortle, J.; Thompson, F.; Harris, C. *Fundamentals of Queueing Theory*; Wiley: New York, NY, USA, 2008. [[CrossRef](#)]
80. Maestre Vidal, J.; Sotelo Monge, M.; Martinez Monterrubio, S. EsPADA: Enhanced Payload Analyzer for malware Detection robust against Adversarial threats. *Future Gener. Comput. Syst.* **2020**, *104*, 159–173. [[CrossRef](#)]
81. Tedesco, G.; Aickelin, U. Strategic Alert Throttling for Intrusion. In Proceedings of the 4th International Conference on Information Security (WSEAS), Tenerife, Spain, 16–18 December 2005; pp. 246–251.
82. Corona, I.; Giacinto, G.; Roli, F. Adversarial attacks against intrusion detection systems: Taxonomy, solutions and open issues. *Inf. Sci.* **2013**, *239*, 201–225. [[CrossRef](#)]
83. Maestre Vidal, J.; Orozco, A.; Villalba, L. Online masquerade detection resistant to mimicry. *Expert Syst. Appl.* **2016**, *61*, 162–180. [[CrossRef](#)]
84. FIRST. *Magerit v.3*; FIRST: Madrid, Spain, 2020.
85. FIRST. *Common Vulnerability Scoring System (CVSS) v3.1*; FIRST: Madrid, Spain, 2020.
86. ENISA. *Methodology for Evaluating Usage and Comparison of Risk Assessment and Risk Management Items*; ENISA: Attiki, Greece, 2020.
87. Llopis, S.; Hingant, J.; Perez, I.; Esteve, M.; Carvajal, F.; Mees, W.; Debatty, T. A comparative analysis of visualisation techniques to achieve cyber situational awareness in the military. In Proceedings of the 2018 International Conference on Military Communications and Information Systems (ICMCIS), Warsaw, Poland, 22–23 May 2018; pp. 1–7.
88. Shameli-Sendi, A.; Louafi, H.; Wenbo, H.; Cheriet, M. Dynamic Optimal Countermeasure Selection for Intrusion Response System. *IEEE Trans. Dependable Secur. Comput.* **2016**, *15*, 755–770. [[CrossRef](#)]
89. Miehling, E.; Rasouli, M.; Teneketzis, D. A POMDP Approach to the Dynamic Defense of Large-Scale Cyber Networks. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 2490–2505. [[CrossRef](#)]
90. Llansó, T.; McNeil, M.; Noteboom, C. Multi-Criteria Selection of Capability-Based Cybersecurity Solutions. In Proceedings of the 52nd Hawaii International Conference on System Sciences, Maui, HI, USA, 8–11 January 2019.
91. Chin, J.; Diehl, V.; Norman, K. Development of an instrument measuring user satisfaction of the human-computer interface. In Proceedings of the SIGCHI conference on Human factors in computing systems, Washington, DC, USA, 15–19 May 1988; pp. 213–218.